THE 8[TH] INTERNATIONAL DEFENCE AND HOMELAND SECURITY SIMULATION WORKSHOP

SEPTEMBER 17 - 19 2018
BUDAPEST, HUNGARY

EDITED BY
AGOSTINO G. BRUZZONE
ROBERT SOTTILARE

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2018

ISBN 978-88-85741-13-3 (Paperback)
ISBN 978-88-85741-12-6 (PDF)

ISBN 978-88-85741-13-3 (Paperback)
ISBN 978-88-85741-12-6 (PDF)

THE 8TH INTERNATIONAL DEFENSE AND HOMELAND SECURITY SIMULATION WORKSHOP, DHSS 2018

SEPTEMBER 17 - 19 2018

BUDAPEST, HUNGARY

ORGANIZED BY

DIME – UNIVERSITY OF GENOA

LIOPHANT SIMULATION

SIMULATION TEAM

IMCS – INTERNATIONAL MEDITERRANEAN & LATIN AMERICAN COUNCIL OF SIMULATION

DIMEG, UNIVERSITY OF CALABRIA

MSC-LES, MODELING & SIMULATION CENTER, LABORATORY OF ENTERPRISE SOLUTIONS

HUNGARIAN ACADEMY OF SCIENCES CENTRE FOR ENERGY RESEARCH

AUTONOMOUS UNIVERSITY OF BARCELONA

MODELING AND SIMULATION CENTER OF EXCELLENCE (MSCOE)

LATVIAN SIMULATION CENTER – RIGA TECHNICAL UNIVERSITY

LOGISIM

LSIS – LABORATOIRE DES SCIENCES DE L'INFORMATION ET DES SYSTEMES

MIMOS – MOVIMENTO ITALIANO MODELLAZIONE E SIMULAZIONE

MITIM PERUGIA CENTER – UNIVERSITY OF PERUGIA

BRASILIAN SIMULATION CENTER, LAMCE-COPPE-UFRJ

MITIM - McLEOD INSTITUTE OF TECHNOLOGY AND INTEROPERABLE MODELING AND SIMULATION – GENOA CENTER

M&SNET - McLEOD MODELING AND SIMULATION NETWORK

LATVIAN SIMULATION SOCIETY

ECOLE SUPERIEURE D'INGENIERIE EN SCIENCES APPLIQUEES

FACULTAD DE CIENCIAS EXACTAS. INGEGNERIA Y AGRIMENSURA

UNIVERSITY OF LA LAGUNA

CIFASIS: CONICET-UNR-UPCAM

INSTICC - INSTITUTE FOR SYSTEMS AND TECHNOLOGIES OF INFORMATION, CONTROL AND COMMUNICATION

NATIONAL RUSSIAN SIMULATION SOCIETY

CEA - IFAC

UNIVERSITY OF BORDEAUX

UNIVERSITY OF CYPRUS

DUTCH BENELUX SIMULATION SOCIETY

I3M 2018 INDUSTRIAL SPONSORS

CAL-TEK SRL

LIOTECH LTD

MAST SRL

SIM-4-FUTURE

I3M 2018 MEDIA PARTNERS

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF SIMULATION AND PROCESS MODELING

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF SERVICE AND COMPUTING ORIENTED MANUFACTURING

IGI GLOBAL – INTERNATIONAL JOURNAL OF PRIVACY AND HEALTH INFORMATION MANAGEMENT (IJPHIM)

HALLDALE MEDIA GROUP: THE MILITARY SIMULATION AND TRAINING MAGAZINE

HALLDALE MEDIA GROUP: THE JOURNAL FOR HEALTHCARE EDUCATION, SIMULATION AND TRAINING

SAGE
SIMULATION TRANSACTION OF SCS

DE GRUYTER
INTERNATIONAL JOURNAL OF FOOD ENGINEERING

MDPI - SUSTAINABILITY

EUROMERCI: THE ITALIAN MONTHLY LOGISTICS JOURNAL

EDITORS

AGOSTINO G. BRUZZONE
MITIM-DIME, UNIVERSITY OF GENOA, ITALY
agostino@itim.unige.it

ROBERT SOTTILARE
US ARMY RESEARCH LABORATORY, USA
Robert.Sottilare@us.army.mil

THE INTERNATIONAL MULTIDISCIPLINARY MODELING AND SIMULATION MULTICONFERENCE, I3M 2018

GENERAL CO-CHAIRS

AGOSTINO G. BRUZZONE, *MITIM DIME, UNIVERSITY OF GENOA, ITALY*
MIQUEL ANGEL PIERA, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*

PROGRAM CO-CHAIRS

FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
YURY MERKURYEV, *RIGA TECHNICAL UNIVERSITY, LATVIA*

THE 8TH INTERNATIONAL DEFENCE AND HOMELAND SECURITY SIMULATION WORKSHOP, DHSS 2018

GENERAL CHAIR

AGOSTINO G. BRUZZONE, ***MITIM-DIME, UNIVERSITY OF GENOA, ITALY***

PROGRAM CHAIR

ROBERT SOTTILARE, ***US** ARMY RESEARCH LABORATORY, USA*

DHSS 2018 INTERNATIONAL PROGRAM COMMITTEE

CAROLE ADAM, LIG – UJF, FRANCE
JONI AMORIM, UNIVERSITY OF SKÖVDE, SWEDEN
MATTEO AGRESTA, *UNIVERSITY OF GENOA, ITALY*
CLAUDIA BAISINI, *SWEDISH NATIONAL DEFENSE COLLEGE, SWEDEN*
MARCO BIAGINI*, NATO M&S COE, ITALY*
JOY BRUCE*, VASTPARK, USA*
AGOSTINO G. BRUZZONE, *UNIVERSITY OF GENOA, ITALY*
RICCARDO DI MATTEO, *UNIVERSITY OF GENOA, ITALY*
JEAN GUY FONTAINE, *NATO STO CMRE*
JOAQUIN GARCIA-ALFARO, *TELECOM SUDPARIS, FRANCE*
PETER KEMPER, COLLEGE OF WILLIAM AND MARY, USA
FRANCESCO LONGO, *UNIVERSITY OF CALABRIA, ITALY*
EDGAR ALONSO LOPEZ-ROJAS, *BLEKINGE INSTITUTE OF TECHNOLOGY, SWEDEN*
MARINA MASSEI, *UNIVERSITY OF GENOA, ITALY*
LETIZIA NICOLETTI*, CAL-TEK SRL, ITALY*
GABRIELE OLIVA, *UNIVERSITY CAMPUS BIOMEDICO OF ROME, ITALY*
ANTONIO PADOVANO*, UNIVERSITY OF CALABRIA, ITALY*
FEDERICA PASCUCCI, *UNIVERSITY OF ROMA 3, ITALY*
ALEXANDRE PETRENKO, CENTRE DE RECHERCHE INFORMATIQUE DE MONTREAL, CANADA
PAOLO PROIETTI, MIMOS, ITALY
ELAINE RAYBOURN, *SANDIA NATIONAL LABORATORIES, USA*
ROBERT SOTTILARE, *ARMY RESEARCH LABORATORY, USA*
RICHARD ZALUSKI, *CSCSS, UK*

TRACKS AND WORKSHOP CHAIRS

PORTS AND LITTORAL PROTECTION
CHAIR: FRANCESCO LONGO, UNIVERSITY OF CALABRIA, ITALY

SERIOUS GAMES: TOWARD EFFECTIVE TRAINING
CHAIRS: PAOLO PROIETTI, MIMOS, ITALY; ELAINE RAYBOURN, SANDIA NATIONAL LABS/ADL, USA

URBAN AREA SECURITY
CHAIR: AGOSTINO G. BRUZZONE, SIMULATION TEAM, UNIVERSITY OF GENOA (ITALY)

CYBERSECURITY
CHAIR: RICHARD ZALUSKI, CSCSS, USA

ADAPTIVE AND PREDICTIVE COMPUTER-BASED TUTORING
CHAIR: ROBERT SOTTILARE, ARMY RESEARCH LABORATORY, USA

NEXT GENERATION C4ISR MODELING & SIMULATION TECHNOLOGY: MIXED REALITY, CROWD SOURCING AND INTEROPERABLE C2-LVC ARCHITECTURES TO SUPPORT OPERATIONS, TRAINING, AND ANALYSIS
CHAIRS: *MARCO BIAGINI, UNIVERSITY OF GENOA, ITALY; JOY BRUCE, VASTPARK*

SIMULATIONS AS TOOLS TO DEVELOP INTUITIVE CONTEXTUAL SENSITIVITY, RE-FRAMING, AND MENTAL AGILITY
CHAIR: CLAUDIA BAISINI, SWEDISH NATIONAL DEFENCE COLLEGE, SWEDEN

SIMULATION FOR AUTONOMOUS & INTELLIGENT SYSTEMS: MOVING FROM COORDINATION TO COLLABORATION AND COMPETITION
CHAIRS: AGOSTINO G. BRUZZONE, UNIVERSITY OF GENOA, ITALY; JEAN GUY FONTAINE, NATO STO CMRE (ITALY)

COMPUTER SIMULATION FOR SECURITY
CHAIRS: EDGAR ALONSO LOPEZ-ROJAS, BLEKINGE INSTITUTE OF TECHNOLOGY, SWEDEN; JONI A. AMORIM, UNIVERSITY OF SKÖVDE, SWEDEN; RICHARD ZALUSKI, CENTER FOR STRATEGIC CYBERSPACE + SECURITY SCIENCE, UK

CRITICAL SYSTEMS MODELING AND SIMULATION
CHAIRS: JOAQUIN GARCIA-ALFARO, TELECOM SUDPARIS, FRANCE; FRANCESCO LONGO, UNIVERSITY OF CALABRIA, ITALY

WELCOME MESSAGE 2018

The International Defense and Homeland Security Simulation Workshop, DHSS, has been established for the first time in 2011 as part of the International Multidisciplinary Modeling & Simulation Multi-Conference (I3M) that was held in Rome. At that time, the opportunity to set-up a specific workshop on Modeling & Simulation for Defense & Homeland Security was also offered by the co-location between the I3M Multi-Conference and the NATO CAX Forum. From that first successful experience, DHSS has (every year) provided experts and scientists with a dedicated workshop to share ideas, to present new methodologies & technologies and to discuss innovative researches.

Selected articles included in the DHSS 2018 proceedings cover different topics ranging from serious games based environments to Virtual Reality and Augmented Reality for training in military domain (but also for crime scene investigation), from cyber security to critical infrastructures protection, from machine learning to intelligent tutoring, from agent based simulation to distributed and interoperable simulation. To this end, particular thanks go to the track chairs, to the members of the International Program Committee and to the Reviewers all for their invaluable work. Furthermore, the DHSS 2018 program also includes joint sessions with the other I3M conferences. This is a traditional aspect of the I3M Multi-Conference, where the idea of thematic conferences (that give attendees the possibility to focus on specific areas of interest) is “augmented” by tracks and sessions cross-fertilization (to create new collaboration opportunities and joint multidisciplinary researches). The DHSS 2018 includes one joint session between DHSS and VARE (the International Conference on Virtual and Augmented Reality in Education) showing the continuous importance of Virtual & Augmented Reality for the Defense and Homeland Security domain and one joint session between DHSS and HMS (the International Conference on Harbor, Maritime and Multimodal Logistics Modeling & Simulation).

Least but not last, a special thanks to all the authors, the success of the DHSS workshop is completely due to their research efforts and to their willing to share results within the DHSS and I3M community.

Again, welcome to DHSS 2018, we hope you will find many useful topics in these proceedings and we wish you a fruitful conference and a pleasant stay in Budapest!

Agostino G. Bruzzone
MITIM-DIME
University of Genoa, Italy

Robert A. Sottilare
U.S. Army Research
Laboratory, USA

ACKNOWLEDGEMENTS

The DHSS 2018 International Program Committee (IPC) has selected the papers for the Conference among many submissions; therefore, based on this effort, a very successful event is expected. The DHSS 2018 IPC would like to thank all the authors as well as the reviewers for their invaluable work.

I3M 2018 INTERNAL STAFF

MATTEO AGRESTA, *SIMULATION TEAM, ITALY*
AGOSTINO G. BRUZZONE, *DIME, UNIVERSITY OF GENOA, ITALY*
ALESSANDRO CHIURCO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
RICCARDO DI MATTEO, *SIMULATION TEAM, ITALY*
JESSICA FRANGELLA, *CAL-TEK SRL, ITALY*
CATERINA FUSTO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
LUCIA GAZZANEO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
MARINA MASSEI, *DIME, UNIVERSITY OF GENOA, ITALY*
LETIZIA NICOLETTI, *CAL-TEK SRL, ITALY*
ANTONIO PADOVANO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
ANTONIO REDA, *CAL-TEK SRL, ITALY*
CATALDO RUSSO, *CAL-TEK SRL, ITALY*
KIRILL SINELSHCHIKOV, *SIMULATION TEAM, ITALY*
CARMINE TOTERA, *CAL-TEK SRL, ITALY*
MARCO VETRANO, *CAL-TEK SRL, ITALY*
BEATRICE ZACCARO, *SIMULATION TEAM, ITALY*

This International Workshop is part of the I3M Multiconference: the Congress leading **Simulation around the World and Along the Years**

Index

STRATAGEMS: AN INNOVATIVE APPROACH FOR INCREASING COGNITIVE AUTHENTICITY IN GAME-BASED ENVIRONMENTS

Emily Newsome[a], Laura Militello, M.A.[b], Sowmya Ramachandran, Ph.D.[c]

[a] ShadowBox LLC, Dayton, OH, USA
[b]Applied Decision Science, Dayton, OH, USA
[c]Stottler Henke Associates, Inc., San Mateo, CA, USA

[a] enewsome@shadowboxtraining.com, [b] l.militello@applieddecisionscience.com, [c]sowmya@stottlerhenke.com

ABSTRACT
Combat Search and Rescue aircrew are tasked with challenging missions, usually under conditions of time pressure, dynamic conditions, and a high degree of uncertainty. Many skills required for successful outcomes (e.g., solving problems quickly; accurately evaluating risks; adapting to rapidly changing environmental conditions) are acquired and maintained through first-hand experiences. This paper describes a research project to develop a pedagogically effective game-based trainer for cognitive skills required for Combat Rescue Helicopter aircrew. Training scenarios are developed using cognitive interviewing techniques in order to promote cognitive authenticity. The trainer will pose critical decisions and provide expert feedback utilizing ShadowBox®, a proven coaching technique that enables trainees to obtain insight into the decision-making processes and reasoning of experts.

Keywords: training, simulation, cognitive skills, game-based training.

1. INTRODUCTION

Personnel Recovery plays a critical role in the success of military missions for both humanitarian and strategic reasons. A chief duty of the United States Air Force is to maintain training of Combat Search and Rescue (CSAR) teams that undertake these important missions. The dynamic nature of these missions requires the ability to be adaptable, solve problems quickly, and maintain a high level of situational awareness in addition to executing tactics and procedures needed for flight in challenging environments.

To prepare aircrew to operate in such complex environments, the Air Force designed a rigorous training program that incorporates a blend of schoolhouse instruction and hands-on training with simulations and live equipment. In the interest of economy and efficiency, simulations are used for most non-tactical training. The Air Force has developed high-fidelity full mission simulators, including custom hardware and software. These simulators typically have a large footprint, and have the benefit of offering a high degree of realism. However, they are often comparable in cost to the equipment being simulated and require a team of dedicated evaluators and technicians to run the simulations. Furthermore, they are not portable; it is costly to move them around, and this limits them to use cases where training is carried out in a fixed location. There are also substantial costs associated with updating them to maintain consistency with real equipment; this is particularly daunting when new equipment such as a new aircraft is introduced. There is a need for training solutions that provide a high level of training effectiveness at a lower cost, with increased portability, and the flexibility required to update training based on evolving platform capabilities and missions.

With regard to training content, flight simulators provide extensive training in tactics, standard operating procedures, and flying skills. Cognitive skills associated with managing the larger mission (e.g., making critical decisions, assessing risk, making sense of the environment, and anticipating next steps) are typically acquired and maintained through mission experience. To complement and extend traditional training for CSAR aircrew, a need was identified for a training platform to build aircrew's cognitive skills. This type of training would be especially useful for overseas deployments, mission rehearsal, and training new missions.

Desktop flight simulations offer a potential strategy for addressing issues of portability and cost-effectiveness while still maintaining the pedagogical benefits of "hands-on" training. Although the degree of realism for manoeuvrability inputs and environmental cues is lessened on desktop simulators, Stewart, Johnson, and Howe (2008) found that a high degree of physical fidelity alone does not necessarily lead to better training outcomes. Instead, criterion-based training strategies are linked with better training outcomes, even when simulators have lower fidelity. A trainer for cognitive skills must have a degree of cognitive fidelity, defined as the extent that the trainer mirrors the actual cognitive activities of the real-world task (Hochmitz and Yuviler-Gavish 2011).

Games, like simulations, are an increasingly attractive and often low-cost alternative to full-mission simulators and have the added benefit of being highly engaging and motivational. In a recent meta-analysis, games were found to lead to 11% higher factual knowledge, 14% higher skill-based knowledge, and 9% higher retention rates in employees who played serious games as part of their training as compared to those who had not played such games (Stizmann 2011). Due to the highly goal-oriented nature of CSAR missions, a game-based trainer is a natural way to train and evaluate cognitive skills of aircrews.

1.1. Stratagems

Our proposed solution is the Stratagems training platform. At its core, Stratagems is a game-based training and mission rehearsal system centred upon cognitively immersive scenarios. Learners are tasked with making high-level decisions about mission planning and execution within the context of rapidly evolving scenarios based on real-world CSAR and MEDEVAC missions. The game-based training setting incentivizes trainees to focus on their own processes and decision-making, and provides opportunities to explore different paths—including those that lead to failure—in a safe environment. Stratagems leverages the pedagogical capabilities of ShadowBox®, a scenario-based training technique that gives learners insight into an expert's thinking during challenging missions (Hintze 2008). Embedded throughout ShadowBox scenarios are cognitive probes, during which learners are required to respond to a question about their current priorities, risk assessment, or assessment of the situation. ShadowBox provides learners with the opportunity to directly compare their answers with the way a subject matter expert (SME) would think about the same situation. Prior to the learner play-through, experts have recorded their feedback about the most and least critical information at specific points in the scenario, how they make sense of complex settings, how they prioritize and establish feasible goals, and the rationale behind their actions. Users also receive a score based on how closely their answers match answers given by SMEs. The ShadowBox approach has shown promise for training cognitive skills in a variety of domains, including with Army and Marine personnel, law enforcement, nurses, social workers, and petrochemical panel operators (Klein and Borders 2016; Flanders, Gunn, Wheeler, Newsome, and Klein 2017; Newsome and Klein 2017; Borders, Polander, Klein, and Wright 2015).

We integrated the ShadowBox functionality of cognitive probing and providing expert feedback to the Stratagems platform. This added capability provided an innovative enhancement, which became the basis for a platform focused on providing training for higher-level cognitive skills in CSAR aircrew, such as planning, decision making, and evaluating risk.

1.2. Overview

This paper will describe the methodological approach for creating an initial prototype of a Stratagems scenario. The technical aspects of the Stratagems game (i.e., game engine, asset selection, non-person characters) are fairly standard for game development and will not be presented. This paper will also discuss our next steps following the completion of the prototype, including gathering user acceptance feedback, pursuing a performance evaluation of the approach, and developing an authoring tool to streamline the scenario generation process.

2. METHOD

Development of Stratagems training included 3 steps: (1) literature review to gain an understanding of the cognitive requirements for CSAR missions; (2) cognitive task analysis (CTA) interviews to elicit critical incidents; and (3) Scenario development, including validation from Subject Matter Experts (SMEs).

First, we began with a literature review, examining publicly available accounts of rescue incidents such as *Leave No Man Behind: The Saga of Combat Search and Rescue* (Galdorisi and Phillips 2009) to understand features that increase complexity in CSAR missions. We also reviewed the Combat Rescue Helicopter Weapons System Training System Requirements Analysis Version 1.0 (TSRA) (Air Combat Command 2015) to identify documented cognitive skill requirements.

Second, we conducted a streamlined cognitive task analysis consisting of interviews with two experienced CSAR pilots and one experienced CSAR gunner/flight engineer. We used the Critical Decision Method (Klein, Calderwood, and MacGregor 1989; Crandall, Klein, and Hoffman 2006) to obtain first-person accounts of challenging, real-world scenarios. The Critical Decision Method is a semi-structured interview technique, during which interviewer and interviewee explore a challenging incident from the interviewee's experience in depth in order to gain a deep understanding of the interviewee's cognitive processes, as well as the context in which the challenging incident took place. Interviews were directed toward cognitive activities such as making sense of a rapidly changing situation, communicating with team members, assessing risk throughout the incident, and identifying the key factors of safety decisions. Although 6-8 interviews is generally recommended when conducting a CTA (Militello and Hutton 1998), because our objective was narrow (i.e., identify challenging incidents) and access to experienced CSAR personnel was limited, we used a streamlined approach in which we explored 1-2 incidents with each of the three interviewees.

Third, we examined findings from the literature review and the streamlined cognitive task analysis to develop a scenario for an initial prototype. Interview notes were transcribed within one week of conducting interviews. We reviewed each incident to determine which provided the best foundation for a first training scenario. We considered the factors that added complexity to determine which could be effectively developed in our initial prototype. We considered whether to focus on the pilot or the flight engineer perspective, and whether the scenario presented challenges that would provide a learning experience linked to cognitive learning objectives. These considerations led us to focus on one incident relayed by one of the pilot interviewees.

After selecting an incident to use as a foundation for the training scenario, we created a decision flow chart providing an overview of the incident, key decisions, and potential variations. Table 1 below provides a section of the decision flow chart for the first decision in the scenario.

Table 1. Excerpt of Decision Flow Chart

	1900: Come on to shift
Information Learned	• Special Operators are doing a mission that requires Rescue crews to stay close. They've started earlier than usual today. • Mission is already underway – you are already 20 minutes behind • Weather: low visibility due to new moon and recent storm that left overcast condition. • Package composition: o 2 HH-60s, 1 AC130, A10s.

Decisions	• What information do you need before you leave?
	  o Why did SOF team leave early?
	  o What is their mission?
	  o Location of enemy threats?
	  o Weather forecast / reports?

Variations	• Weather
	  o Foggy, sandstorm
	  o Daylight, clear night
	• How far behind is player? (sense of urgency)
	• Crew composition
	  o New people?
	  o 1 or 2 SMAs?
	  o PJs? Do you know them? Are they experienced?

Using this as a frame, we drafted a narrative timeline, and cognitive probes intended to encourage the learner to articulate key concerns, priorities, and actions at critical points in the scenario. We asked the two CSAR pilots who had participated in interviews to review the narrative timeline, and to provide a response and rationale statement for each cognitive probe. We refined the scenario based on this input, and drafted an expert model linked to each cognitive probe. An experienced pilot who did not participate in the cognitive interviews reviewed the refined scenario, and provided additional feedback and elaboration on the expert model.

Basing our scenario development on CTA findings allowed us to incorporate a high level of cognitive fidelity into the scenario design process. The literature review added breadth to our small-sample CTA, and also helped us stay grounded in training requirements as defined by the U.S. Air Force. Direct feedback from experts about their decision making is an important component of the pedagogical approach of the Stratagems system. The integrated rankings and rationale from multiple SMEs, provided the expert response model to which learner responses will be compared during training.

3. RESULTS

Results from the literature review, the cognitive task analysis, and the resulting training scenario are described in turn.

3.1. Literature review

3.1.1. Common decisions and cognitive skills

To identify a candidate set of critical cognitive skills that could be addressed with Stratagems, we reviewed the Combat Rescue Helicopter Training Requirements Analysis (TSRA) Version 1.1 Master Task Training List (MTTL) (Air Combat Command 2015). Specifically, we examined the TSRA for tasks that require judgment and problem solving. We organized critical skills into a list of skills of interest during Cognitive Task Analysis interviews. Skills were organized into the following categories based on macrocognitive functions: communication, crew management/delegation, sensemaking, anticipating/planning, and monitoring safety.

Table 2. CSAR Aircrew Cognitive Skills

Cognitive Skills
Communication
Communicate with other crewmembers to obtain/relay/provide mission details using tactical communications systems
Brief crewmembers before, during, and after missions—presenting info in logical sequence
Crew management
Manage resources and duties effectively to minimize task saturation, channelized attention, and distractions among crewmembers
Direct maintenance support to correct discrepancies noted during pre-flight
Sensemaking
Understand changing situations and environments
Determine if any restrictions exist on departure, en-route, and at destination
Perform scanning during all flight operations
Analyse weather briefing
Perform battle damage assessment
Recognize ground and airborne threats to aircraft
Anticipating/planning
Evaluate feasibility of divert
Prepare alternate plan
Perform inflight dynamic planning
Perform Landing Zone (LZ) options
Anticipate and respond to terminal area contingencies

Cognitive Skills
Monitoring safety
Apply sound judgment with regard to mission accomplishment and safety
Perform survivor authentication
Maintain aircraft parameters within limits of EM charts

Interviewers were primed to explore the MTTL critical skills as they arose, but were also open to exploring decisions and tasks not previously identified. In addition to identifying specific skills from the MTTL, we reviewed the rescue incidents described in Galdorisi and Phillips' (2009) *Leave No Man Behind: The Saga of Combat Search and Rescue* to understand the operational context of CSAR pilots. We used the incidents to explore how critical skills in the MTTL might play out in real-world scenarios. Some examples of situated critical cognitive skills from our analysis of incidents in the book include:

- Determine whether an area is "too hot" to attempt a rescue.
- Determine the rescue strategy (e.g., landing the helicopter versus hovering to allow personnel to rappel down).
- Modify the mission to deal with unforeseen developments (e.g., more enemy fire than anticipated, weather changes, low fuel due to not being able to locate person).
- React to observations from back crew (e.g., "I see power lines at 900, climb, climb, climb!").

Leave No Man Behind (2009) presented recollections of CSAR missions from conflict in the jungles of Vietnam in the 1960s and 1970s to more recent conflict in Afghanistan. Through a review of these incidents, we also identified features that make CSAR missions especially challenging. These include:

Table 3: Features that increase complexity in rescue scenarios

Features that increase complexity
• Uneven terrain (complicates finding a suitable place to hover or to land) • Technical difficulties (i.e., winch/rappel line won't work) • Injured personnel that are immobile or severely injured • Rescuing multiple people (i.e., added weight, accountability) • Enemy fire • Not being able to locate the pilot (faulty Global Positioning System (GPS), etc.) • Weather changes (i.e., thunderstorm rolling in, fog) • Environmental challenges (i.e., high altitude, high temperature) • Urban environment (i.e., don't know where enemy is and how many there are) • Low fuel

Taken altogether, these three lists (list of cognitive skills in MTTL, situated cognitive skills, and features that increase complexity) informed the general roadmap for the types of incidents we wanted to explore during CTA interviews. Although interviews were geared toward the concepts identified in the literature review, we remained open to discoveries and interesting problems interviewees had encountered in their careers.

3.2. Cognitive Task Analysis

CTA interviews resulted in three incidents—one from each interviewee. Only one incident was a true CSAR mission, defined by being carried out in or near combat zones during a time of war. The other two incidents included a Medical Evacuation (MEDEVAC) mission, and an attempted but aborted rescue mission. Each incident is described in detail below.

3.2.1. Wreck in the ROZ.

This pilot was commanding CHALK 2. The rescue team was tasked with providing back up support for Special Operations Forces (SOF) as they conducted combat missions in Iraq. On this day when the pilot arrived on shift, he learned that the SOF had scrambled to a combat zone approximately 100 miles away, 20 minutes ahead of schedule. He and the rescue crew quickly took off and headed toward an initial rendezvous point. En route, they learned that a U.S. Army HH-60 had crashed inside a Restricted Operating Zone (ROZ), which had been set up for the SOF mission. Soon after, the helicopter crew learned that personnel injured in the crash had been ex-filled, but survivors and deceased crewmembers remained at the crash site. The rescue helicopters flew a racetrack holding pattern just outside the ROZ as they attempted to get more information about the crash from both the ROZ owner and the Tactical Operations Centre (TOC). The crew had to decide at each stage whether to proceed with limited information, and whether to violate protocol by entering the ROZ without permission.

3.2.2. High, hot, and heavy on the side of a mountain

This pilot was commanding CHALK 2. Two rescue helicopters and crew had flown south to a nearby base to join the Army for a steak dinner. On the way back from dinner, they saw traffic on their Internet relay chat about a helicopter that had crashed further north. The rescue helicopters offered to go to the crash site to investigate. They decided to land at a nearby base to collect more information before heading to the crash site. They learned that a CH47 (Chinook) was ex-filling an Army special operations team that had been operating in the area. The Chinook should have returned, but no one had heard from them. The most likely explanation was a crash. They knew there had been three crewmembers on the Chinook, but they did

ISBN 978-88-85741-12-6; Bruzzone and Sottilare Eds.

not know the status of the crash or how many of the people being ex-filled had entered the Chinook before the crash. They knew that the terrain near the crash site would be challenging, because it was located on the side of a mountain. Furthermore, the extremely hot weather, the high altitude, and the weight of the survivors would reduce the power of the aircraft. As they approached, they saw the valley below the rendezvous point aflame. The Chinook had struck the side of the mountain when landing, and tumbled down the mountain. During the descent, the Chinook sprayed the valley with jet fuel. The valley was on fire. It was so bright that Night Vision Goggles (NVGs) would no longer work. It was difficult to see much beyond the fire. The crew had to decide whether they could safely attempt a rescue of the surviving Army personnel, and if not, what an alternate plan of action would be.

3.2.3. Popeye in a Sandstorm

This incident was told from the perspective of an SMA. The rescue team was sitting alert when they received the call to scramble. The SMA was doing engine checks, putting on armour, and listening to the radio for the 9-line all at the same time. He learned that there had been an improvised explosive device (IED) explosion, and the injured person had lost both legs and an arm. The injured person was a member of NATO ally forces. This was all happening during an on-going sandstorm. The weather did not look impossible, but the SMA knew the conditions were challenging. As the helicopters took off, within seconds they called blind. The crew had to decide how to reorient and find straight and level flight, then had to formulate a plan to continue with the mission or return. Note: although he didn't know it at the time, the interviewee told us that he later learned a NATO ally MEDEVAC team declined the mission because the weather was too severe.

After careful analysis, we selected the "Wreck in the ROZ" scenario for further development into a prototype scenario based on three criteria. One, the focus on the pilot/co-pilot role was most appropriate for our initial effort, because of the executive decision-making role of pilots/co-pilots. While the SMAs handle the logistics of the mission, the pilot and co-pilot manage executive decision-making, such as making go/no go decisions, selecting routes, navigating the aircraft, receiving and interpreting intelligence briefings, and responding to changes in the environment (e.g., weather, terrain, enemy fire, etc.). Thus, a scenario from the pilot/co-pilot scenario would more directly address the cognitive training objectives articulated in the literature review. Two, the "Wreck in the ROZ" scenario would require limited terrain mapping and therefore would be suitable for a limited prototyping effort. Three, this scenario also required the pilot to make a series of difficult judgments in the face of uncertainty and introduced a range of complex elements, including some from the list of features that increase the complexity of CSAR missions presented above.

3.3. Training scenario

After selecting the *Wreck in the ROZ* incident, we deepened the scenario and added detail, which included integrating ShadowBox probes, articulating learning objectives, and soliciting feedback from Subject Matter Experts. We describe the resulting prototype below.

3.3.1. Gameplay

Because the objective of the Stratagems game is to develop learners' cognitive skills rather than their flying or procedural skills, we deliberately simplified the look and feel of the game environment in order not to detract from the cognitive focus of the game. Thus, helicopter gauges and terrain features were relatively low-fidelity representations. We further emphasized decision making and cognitive skills by removing the need for the learner to pilot the aircraft during the training scenario. This style of gameplay results in an experience that is similar to a flight lead's role in the mission; the duty of flying the aircraft is typically assigned to the co-pilot, while the pilot concentrates on higher-level decision making and communication tasks. Instead of having an active flying role, the learner is tasked with monitoring the flight of the aircraft and to make sense of incoming information. Learners must also respond to ShadowBox probes, when prompted. The choices and rankings learners submit at each decision influence the course of the scenario, as well as the probability of events happening throughout the scenario. For example, selecting "Option A" may result in higher risk of enemy fire because it makes the aircraft more visible to the enemy than "Option B." Figures 1a and 1b provide screenshots of the Stratagems gameplay.

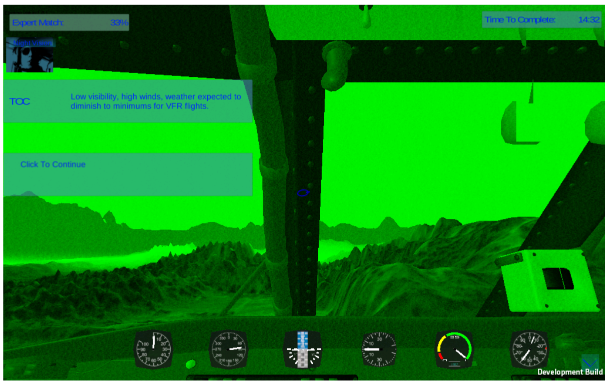

Figure 1a: Stratagems Player View

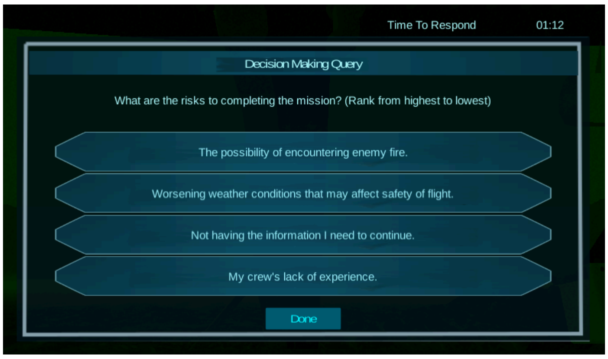

Figure 1b: ShadowBox Probe

3.3.2. ShadowBox Probes

The purpose of the ShadowBox probes is twofold: First, to encourage learners to articulate their understanding and priorities at key points in the scenario, and second, to show the learners how experts think about the same situations. The ShadowBox probes are provided in real-time as the learner plays through a game. Questions asked generally fall into one of three categories: Information needs probes ("What information do you need before leaving?"), risk assessment probes ("What are the risks to completing the mission at this time?"), prioritization probes ("What are your most pressing priorities?"), or strategy probes ("What strategy would you use to land safely?"). Generally, decisions immediately followed an information inject (e.g., updated weather briefing, additional information from the command centre) that changed the learner's understanding of the situation.

As the scenario was instantiated in the game format, we found that we needed a visual representation of the incident to facilitate discussion and common ground across subject matter experts, software developers, and cognitive task analysts. The resulting scenario map (see Figure 2 below) allowed us to quickly orient to specific points in the scenario, consider placement of branches and probes, and quickly identify information gaps and inconsistencies.

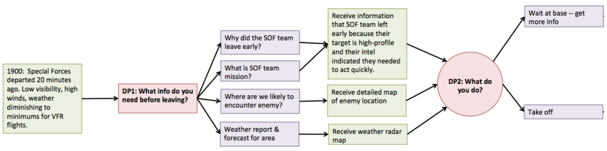

Figure 2: Scenario Branching Diagram

A second critical aspect of the ShadowBox approach is the ability to review an expert's perspective on the scenario and specific decision points. To encourage reflection without disrupting the flow of the game, expert feedback is provided at the end of the scenario. This also gives learners the opportunity to make decisions based on their own assessments and choose actions without being influenced by expert rankings. After completing a scenario, the after-action review (AAR) provides the opportunity for students to compare their rankings with those of experts and reflect on the similarities and differences.

3.4. Scoring

To score performance on Stratagems scenarios, learners will be evaluated using two separate criteria. One measure of performance will be the degree to which their chosen tactics, strategies, risk perception and decision making capabilities align with the expert model at decision moments. Learners will receive an overall percentage match score, and will be encouraged to review decisions in which their answers did not align with SME answers. Due to the heavy emphasis on outcomes in this domain, we wanted to increase the realism of our scoring paradigm by rewarding learners for not only making the right choices, but also for achieving the most desirable outcomes. Therefore, a second performance measure will be based on the scenario outcome attained based on the learner's chosen strategy (e.g. if injured personnel were rescued; if the helicopter was damaged or crashed; if rules of engagement or protocol were violated).

4. SUMMARY AND NEXT STEPS

The focus of Stratagems is to create a game-based trainer with a high degree of cognitive authenticity in order to facilitate the development of cognitive skills. Strategic design choices including the integration of ShadowBox and using Cognitive Task Analysis to inform scenario development, allowed us to address specific learning objectives such as sensemaking, anticipating, and crew management. Decision feedback from experts was incorporated into the scoring function, to encourage learners to compare their choices to those of a panel of experts.

A next step for enhancing the Stratagems tool is to create an authoring tool to allow trainers to create scenarios on their own. Although we believe Cognitive Task Analysis is the best way to create a detailed and challenging scenario, it can often be a time- and resource-consuming process. Therefore, we are designing an authoring tool interface to support instructor pilots in building cognitively challenging scenarios based on their own lived experiences. The system would prompt the trainer to select from certain pre-sets such as terrain, crew, or mission type (e.g., CSAR or MEDEVAC) before helping them build a timeline of the incident and populate decision points. Finally, the system would prompt the trainer to review alternate paths or branches through the scenario and the consequences for choosing alternate decision options.

Another logical next step would be to build a training scenario from the perspective of the SMA. This would require additional literature reviews and cognitive task analysis for examples of cognitive skills and challenging incidents from the SMA perspective. We are also in the process of gathering user acceptance data from current CSAR pilots and trainers. We hope to collect reactions in order to more formally assess the validity of our scenario and decisions, the level of cognitive realism, and the perceived value of the Stratagems training platform.

ACKNOWLEDGMENTS
This material is based upon work supported by the United States Air Force under Contract No. FA8650-17-C-6865. We would like to thank Coray Seifert and K.C. Pake for their contributions to this effort.

REFERENCES

Air Combat Command (ACC)/A5R, 2015. Combat rescue helicopter (CRH) weapon system training system requirements analysis (TSRA). Version 1.0. Wright-Patterson Air Force Base, OH.

Borders, J., Polander, N., Klein, G., and Wright, C., 2015. ShadowBox™: Flexible training to impart the expert mindset. Procedia Manufacturing, 3, 1574-1579.

Crandall, B., Klein, G.A., and Hoffman, R.R., 2006. Working minds: A practitioner's guide to cognitive task analysis. Cambridge: MIT Press.

Flanders, S.A., Gunn, S., Wheeler, M., Newsome, E., and Klein, H.A., 2017. Accelerating the development of higher-level clinical thinking in novice nurses. Journal for Nurses in Professional Development, 33 (5), 240-246.

Galdorisi, G., and Phillips, T., 2009. Leave no man behind: The saga of combat search and rescue. Minneapolis: Zenith Press.

Hintze, N.R., 2008. First responder problem solving and decision making in today's asymmetrical environment. Thesis (Master's). Naval Postgraduate School.

Hochmitz, I., and Yuviler-Gavish, N., 2011. Physical fidelity versus cognitive fidelity training in procedural skills acquisition. Human Factors, 53 (5), 489-501.

Klein, G., and Borders, J., 2016. The ShadowBox approach to cognitive skills training: An empirical evaluation. Journal of Cognitive Engineering and Decision Making, 10 (3), 268-280.

Klein, G.A., Calderwood, R., and Macgregor, D., 1989. Critical decision method for eliciting knowledge. IEEE Transactions on Systems, Man, and Cybernetics, 19 (3), 462-472.

Militello, L.G., and Hutton, R.J., 1998. Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. Ergonomics, 41 (11), 1618-1641.

Newsome, E., and Klein, G., 2017. Cognitive skills training for frontline social workers: A pilot study. Proceedings of the 13th International Conference on Naturalistic Decision Making, pp. 247-251. June 20-23, Bath (UK).

Stewart, J.E., Johnson, D.M., and Howse, W.R., 2008. Fidelity requirements for Army aviation training devices: Issues and answers. Fort Rucker, AL: Army Research Institute for the Behavioral and Social Sciences.

Stizmann, T., 2011. A meta-analytic examination of the instructional effectiveness of computer-based simulation games. Personnel Psychology, 64 (2), 489-528.

AUTHOR BIOGRAPHIES
Emily Newsome is a Research Associate at ShadowBox LLC, where she has investigated the cognitive elements of expertise for military personnel, child welfare workers, and nurses. She has developed training programs utilizing the ShadowBox approach and assisted in implementing and evaluating ShadowBox training across the United States. She earned her Bachelor of Science in Psychology from Wayne State University (Detroit, MI) in 2014.

Laura Militello, MA is a Senior Researcher at ShadowBox, LLC and CEO of Applied Decision Science, LLC. She applies cognitive systems engineering to the design of technology and training to support decision making in complex environments. She focuses primarily on supporting decision making in healthcare and military domains. She is also acknowledged as one of the masters of advanced Cognitive Task Analysis methods, and one of the leaders of the Naturalistic Decision Making movement.

Sowmya Ramachandran, PhD is a Research Scientist and Project Manager at Stottler Henke Associates, Inc. At Stottler Henke, her focus has been on the applications of AI to improve education and training. She has been a principal investigator on a number of efforts to develop Intelligent Tutoring Systems (ITSs) for a variety of domains, including reading comprehension, helicopter piloting, emergency first response, and information technology troubleshooting.
Dr. Ramachandran received her Ph.D. in Artificial Intelligence from The University of Texas at Austin.

CRIME SCENE EVIDENCE COLLECTION AS A VIRTUAL REALITY USE CASE

Jeffrey Mills[(a)], Tabitha Dwyer[(b)], Jennie Ablanedo[(c)], Tamara Griffith[(d)], Elwin Collins[(e)], Ron Acker[(f)]

[(a,c)] University of Central Florida, Institute for Simulation and Training
[(b)] Cole Engineering Services, Incorporated
[(d)] U.S. Army Research Laboratory Simulation & Training Technology Center [(e,f)] Federal Law Enforcement Training Centers

[(a)] jmills@ist.ucf.edu, [(b)] tabitha.dwyer@cesicorp.com, [(c)] jablaned@ist.ucf.edu,
[(d)] tamara.s.griffith.civ@mail.mil, [(e)] elwin.collins@fletc.dhs.gov, [(f)] ronald.p.acker@fletc.dhs.gov

ABSTRACT
Virtual Reality (VR) is an interface strategy that completely immerses the user into a virtual environment. Users with head mounted displays (HMD) see and hear the immersive space as if they are no longer in the real world. Instead they are fully wrapped in the surroundings of the virtual space. While VR has been around for many years, recent advances in commercial technology have provided a surge of cost-effective devices to mainstream end-users. With this surge in hardware availability, software developers have been exploring ways to make use of VR to design engaging training experiences. This paper describes the creation of a VR simulation to train the investigative mindset relative to evidence collection tasks at a crime scene. Spatial orientation within a crime scene is critical to understand blood splatter, munitions trajectory and to conceptually re-create events in the crime scene. This paper will explain the challenges and solutions of developing this VR experience. Finally, the paper will provide lessons learned and recommended future development involving this VR training application.

Keywords: Virtual Reality, Crime Scene Evidence Collection, Serious Games, Training Games, Homeland Security, Virtual Environments

1. INTRODUCTION
Virtual Reality (VR) represents the simulation of reality. A virtual environment is a three-dimensional representation of a computer generated space (Maxwell, Griffith, & Finkelstein, 2014), but VR fully immerses a person in that space. Gigante (2014) describes VR as:

> "The illusion of participation *in* a synthetic environment rather than external observation of such an environment. VR relies on three-dimensional (3D), stereoscopic, head-tracked displays, hand/body tracking and binaural sound. VR is an immersive multi-sensory experience."

VR headsets gained popularity with launches of the Oculus Rift, HTC Vive and Sony HMZ in 2016, which were priced for the masses, but the technology has still not gained widespread acceptance by mainstream society (Krewell, 2017). It appears that industry is still exploring various user interfaces and seeking where the technology best fits.

In a similar way, the military and first responders are also exploring the technology to understand how it can support training tasks. One such area of research is focused on training various first responders and is described in this paper.

The Federal Law Enforcement Training Center (FLETC) provides career-long training to law enforcement professionals to fulfill their responsibilities safely and proficiently. FLETC has grown into the Nation's largest provider of law enforcement training. Under a collaborative model, FLETC's federal partner organizations provide specialized programs to meet their mission goals. This includes areas common to all law enforcement officers such as training on fire arms, tactics, driving and investigations to name a few (Federal Law Enforcement Training Center, 2015). The organization incorporates various forms of simulation to augment their programs, including augmenting live scenarios with virtual simulation to gain efficiency and effectiveness.

FLETC partnered with the Army Research Lab to explore the possibilities of the growing use of VR for training. The Army Research Lab has a long history of contributions toward advancing training capabilities to support soldier training. The partnership allowed each team to contribute their expertise toward researching the bounds of VR for certain training tasks. This research conducted under this partnership benefits both first-responders and soldiers as well as contributing to the knowledge base for anyone interested in making use of VR. The team elected to build a prototype that makes use of the Enhanced Dynamic Geo-Social Environment (EDGE). EDGE is a Government owned platform built on the Unreal 3 and 4 commercial game engine. It is used to support various training tasks as well as to research the use of commercial game technology in support of training tasks shared across Government agencies. Leveraging EDGE provided FLETC with a low-cost government furnished software solution with access to source code making it possible to accelerate the development timeline. The development team at the U.S. Army Research Laboratory (ARL), Simulation & Training Technology Center (STTC) provided FLETC with a playable prototype of a crime scene evidence collection simulation in Virtual Reality. This prototype can answer questions such as:

- Is the user more comfortable and engaged in VR or is the use of desktop computers with a keyboard and mouse preferable?
- Could this prototype be used as a practical exercise after teacher-led training?
- Does this technology increase accessibility as compared to using a physical live training area?
- Can this technology reduce space requirements, reduce setup time and reduce trainee downtime as compared to live training exercises?

This paper describes the process of developing the prototype, the lessons learned that can be applied to other training tasks, and provides recommended paths forward in relation to the use of VR for training certain tasks.

2. VIRTUAL REALITY PROTOTYPE CREATION

VR is a platform that can focus trainees and allow for a powerful message to be delivered through a distraction-free training technique. Researchers (Jang, Vitale, Jyung, & Black, 2017) have demonstrated that direct interaction in VR of spatial tasks facilitates embodiment of those tasks and helps the learner maintain a clear frame of reference for later recall.

The applicability of using VR in a spatial training task was the exploratory research topic. A VR prototype of a crime scene was created. The first step was to conduct a stakeholder's requirements discussion. Next, an initial design would be presented to the stakeholders for feedback. After stakeholder buy-in, art and software coding was conducted in a spiral manner to welcome continuous stakeholder feedback. Then final testing and debugging followed. The development team remained in contact with FLETC throughout this process to provide progress reports and course changes as necessary. The team's initial design ideas evolved over time due to feedback or changes in customer priorities.

2.1. REQUIREMENTS

The team used the HTC Vive VR headset. The Unreal Marketplace had VR products that provided the development team an excellent starting point. The team started by using an existing, free apartment living room level and an Unreal Engine 4 VR template. The template had the basic pick up and drop functions that work seamlessly with the Vive hand controllers. This allowed the team to build content in a matter of weeks with the bulk of the time used to create assets to specifically apply to a crime scenario. This included bedroom items, an additional bathroom area and both evidence and distracter props. A functional room-wide black light mode was created to show evidence stains and a custom user interface was designed for collection actions. A custom user interface with asset icons was also added for scenario assessment.

During development the team produced object interaction and collection capabilities within the level; this included user interface interactions, controller button responses, character locomotion and task-specific functions. The team made it possible to move about the virtual space, as well as inspect and remove blankets and bed sheets. A black-light functionality was a special function that allows the user to gain additional information to criminal activity.

2.2. USING THE TRAINING SCENARIO

When designing a level within a game engine, it's important to understand the principles of game design. Some important principles considered here include; visual style and game play (Bates, 2004). Visual styles represent how the game appears to the player including the environment, characters and from what perspective the environment will be seen. Gameplay considered ways to engage the user and progress the story using controls or interactions, experience duration, rewards and difficulty (Bates, 2004). Developing for a VR System follows these principles within a 360 degree environment while maintaining frame rates over 90 Hz to avoid simulator sickness (Ahn, Kim, & Kim, 2018).

The team's result is a playable virtual reality training simulation. It enables users to assess the crime scene within a virtual hotel room and carefully inspect and collect evidence. Game play consists of walking within the boundaries of the Vive room setup. The user may also use the handheld controllers to teleport to spots inside of the navigable area. Play is freeform and the user is expected to look on top of surfaces, under and behind items and on the floor. The Vive hand controller allows the collection of evidence, inspection and then choosing whether to place the object into an inventory (evidence collection). All evidence and distractors can be picked up and dropped if not collected. The items collected may be viewed at any time via a button on the controller (See Figure 5). Collected items are displayed through a Graphical User Interface (GUI) window that toggles on and off. The black light mode will expose hidden stains on surfaces. This is also toggled on and off through a controller keystroke. The User may pull the sheets back and forth on the bed to see underneath and gather them if needed. A custom navigation layout was created for the user's movement within the environment. The environment was scaled slightly larger than 1:1 to invoke a more natural experience for VR.

2.2.1. SCENARIO OUTCOME

FLETC was interested in using the crime scene scenario to better understand VR technology for learning and teaching. This project was used as a way to assess the state of the technology and what training tasks the technology lends itself to.

VR is still a relatively immature technology. While Unreal 4 has a vast development community, the pool of developers building VR functionality is a fraction of that for traditional game modalities. Many of the challenges the team experienced have not yet been addressed by the commercial game industry. This will be discussed further in Section 5, Issues and Lessons Learned.

2.2.2. ENVIRONMENT

The environment consists of one game level reproducing a simple hotel room (See Figure 1) and bathroom (See Figure 2).

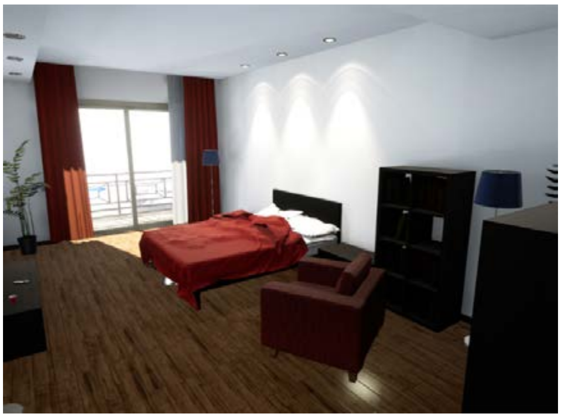

Figure 1: Bedroom

By utilizing an existing game level and a VR plugin from the Unreal Marketplace, the team was able to focus development on evidence asset creation. The VR template had functions to pick up and drop objects using the Vive hand controllers. The team concentrated their effort on 3D evidence models and a custom inventory collection interface. This was a tremendous cost saving to the customer.

Figure 2: Bathroom

A simple drag and drop feature was needed to modify the placement of objects in the environment. All button interactions were mapped for the collection interface window (See Figure 3). This displays the evidence collected. It can be visibly toggled on and off. Names for the objects were added for clarity.

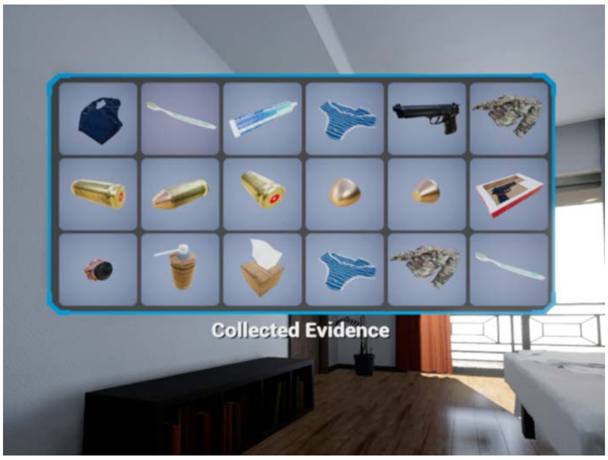

Figure 3: Collection Interface

3. GAME CONTROLS AND USER INTERFACE

The HTC Vive Head Mounted Display (See Figure 4) was the hardware or physical interface used in the prototype.

Figure 4: Head Mounted Display (HTC VIVE, 2011-2018)

The Unreal Engine has direct controller support built in with the HTC Vive plugin (Epic Games, 2016). Mapping all the positions on the controller to interactions in the game is done inside the editor with little difficulty. The controllers for the Vive (See Figure 5) allow the user to take hold of objects within the environment and to navigate. The team used trial-and-error to ergonomically place the most used commands to be intuitive to the player.

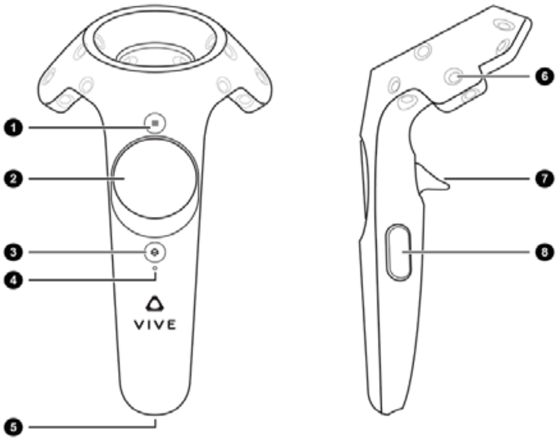

Figure 5: HTC Vive Hand Controller (HTC VIVE, 2011-2018)

4. CUSTOMER DELIVERY

The entire prototype scenario software was delivered to FLETC, including operating instructions. The project contains all art assets, all programming blueprints, code and the complete Unreal Engine 4 editor. This allows the customer to modify the level based on training needs. Simple adjustments do not depend upon a team of programmers and artists. Rearranging evidence items within the rooms and reconfiguring controller button mapping is done directly in the editor by the customer.

5. ISSUES AND LESSONS LEARNED

VR is a relatively new development platform (Earnshaw & Gigante, 2014). The Unreal engine has just recently added the HTC Vive functionality and a template for rapid use (Epic Games, 2016). Development heavily relied on the Unreal Engine community. Currently there is less available support and a minimum of game variety within VR. The challenges we confront (See Section 5.1) with this medium have yet to be adequately addressed by the commercial game community.

5.1. CHALLENGES

Although the project included documentation for best practices from FLETC, subject matter expertise (SME) was purposefully restricted to the FLETC Training Innovation Division personnel with previous law enforcement experience. This decision was made by the FLETC to limit the impact on the training instructor schedules. The FLETC acknowledges that having multiple SMEs heavily involved in the development stages would reduce the risk of missing important aspects of the law enforcement evidence collection process. The prototype provides the appropriate stimulus to support specific training tasks, but with minimal pedagogy, the focus was primarily dedicated to the development of technical capabilities.

5.1.1. User Interface

One of the considerable challenges faced was constructing a user interface that displays properly in VR space. User controls and interface integration proved to be a difficult task because of the 360 degree virtual environment. Decisions upon where GUIs such as the evidence locker would appear may influence whether a person can navigate while the locker is showing or not. The Unreal Engine has motion graphic tools for typical computer game design (Epic Games, 2018). Developing user interfaces in VR is *not* parallel to two dimensional user interfaces. In order to have the user interface display properly, the graphic approach must be rendered in 3D space instead of a typical 2D interface overlay. This means that the interface is seen from the person's point of view regardless of what is being viewed in the environment. In addition, visual resolution for the user's focal length needed to be experimented with to find a good balance.

Another interface issue was with the controller mapping scheme. The initial plan for interaction was not possible within the timeframe allotted. The input for black light, collection, teleporting and viewing the inventory was accomplished without the need to use a graphic menu, but rather through keypad input. Future iterations of the prototype might include a GUI that reminds users what the controller functions are to accomplish these tasks.

5.1.2. Game Physics

The ability to accurately model physics within the virtual environment constrained game play. This is a technology driven constraint that should be resolved with continuing development and popularity of VR systems. The Unreal Engine 4 physics model has a set gravity and mass (Epic, 2018). All objects large or small react physically the same. While this is believable with things like a cup or toothbrush, large objects lack proper mass therefore look unrealistic if moved. For example, a couch lifts as easily as a cup. In addition, some objects tend to bounce violently when dropped and in certain cases would move outside the playable area entirely. Virtual Reality provides individuals with stunningly realistic environments however, immersion is often broken when objects act differently from what is expected. Presently the inability for objects to physically stop a player from walking through chairs and other furniture, or hands passing through things is a limiting factor for this medium.

There were art asset issues concerning cloth physics in VR as well. Part of the crime scenario involves clothing that is lying on the floor. The intent was to have the clothes and bed sheets move realistically when picked up. A limitation was discovered in picking up clothing. Cloth physics require an anchor mesh to attach to. Cloth physics also do not collide with objects in the virtual space. The result is clothing falling through things in the environment if picked up and dropped. The solution was creating clothing as a 3D mesh for interaction however it is visually stiff resembling cardboard.

6. FUTURE DEVELOPMENT

There are many ways to improve and expand the initial prototype. More time spent upfront concerning the issues discovered during development will improve the user experience. An enhanced VR controller based interface system, and collection method as well as a way to indicate virtual object weight would enhance the experience. Also, the inclusion of a VR inverse kinematic system will enhance realistic motion if characters are included. Currently, the inventory collection process has two actions, collect or don't collect. In the future the ability to choose the type of container to store items and cyanoacrylate fuming for fingerprints would be useful features. The addition of photographs and evidence marking tents can be added depending on customer desires.

The most significant feature in a future iteration is a scenario creation and editing tool for the customer. The VR crime scene investigation is a prime candidate for this feature. The ability for FLETC to modify the scene to their needs will expand this tool's use beyond the prototype.

There are further questions to explore. Data collection within the classroom is crucial to understanding the end user's comfort level and engagement utilizing VR. Does VR have an advantage over the use of desktop computers with a keyboard and mouse in a spatial activity such as processing a crime scene?

7. CONCLUSION

Virtual Reality continues to grow in popularity and availability. The entertainment industry is not the only domain. The simulation and training industry sees strong potential for growth, and prototypes like the crime scene scenario is just one example. The spatial elements involved in processing the crime scene make use of strengths that VR brings to the table. Not all situations are ripe for VR. The challenge is identifying the training that can best leverage from the experience. More research in this area is needed to expand and refine the experience for the future of this technology. The FLETC VR crime scene investigation simulation is a step in that direction.

8. REFERENCES

Ahn, J., Kim, Y. Y., & Kim, R. Y. (2018). Virtual Reality-Wireless Local Area Network: Wireless Connection-Oriented Virtual Reality Architecture for Next-Generation Virtual Reality Devices. *Applied Sciences*, 8.

Bates, B. (2004). *Game Design*. Boston, MA: Premier Press.

Earnshaw, R. A., & Gigante, M. A. (2014). Virtual Reality: Definitions, History and Applications. In R. A. Earnshaw, *Virtual Reality Systems* (p. 3). London: Academic Press.

Epic. (2018, January 1). *Unreal Engine 4 Documentation*. Retrieved from Epic Unreal 4.

Epic Games. (2016). HTC Vive Template. Retrieved May 10, 2018, from https://forums.unrealengine.com/development-discussion/vr-ar-development/80368-htc-vive-template

Epic Games. (2018). *Unreal Engine Documentation*. Retrieved from Unreal Motion Graphics: https://docs.unrealengine.com/en-us/Engine/UMG/HowTo/Creating Widgets

Federal Law Enforcement Training Center. (2015). *Learn About FLETC*. Retrieved from Federal Law Enforcement Training Center: http://fletc.gov/learn-about-fletc

HTC VIVE. (2011-2018). VIVE. Retrieved May 14, 2018, from https://www.vive.com

Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education*, 150-165.

Krewell, K. (2017, July 7). The State of the Virtual Reality Business. *Forbes*. Retrieved May 12, 2018, from https://www.forbes.com/sites/tiriasresearch/2017/07/07/the-state-of-the-virtual-reality-business/#2ad3c104dfbe

Maxwell, D., Griffith, T., & Finkelstein, N. l. (2014). The Use of Virtual Worlds in the Military Services as Part of a Blended Learning Strategy. In K. S. Hale, & K. M. Stanney, *Handbook of Virtual Environments: Design, Implementation, and Applications, Second Edition* (pp. 959-999). Orlando, Florida: CRC Press.

ACKNOWLEDGEMENTS

We would like to express appreciation to the Federal Law Enforcement Training Center (FLETC) for their valuable contribution and suggestions during the planning and development of this research work. We would also like to thank Cole Engineering Services Inc. (CESI Corp) for their development efforts.

AUTHORS BIOGRAPHY

Jeffrey Mills

Jeffrey Mills is a Faculty Associate for the University of Central Florida Institute for Simulation and Training (UCF IST). He supports content creation for the Enhanced Dynamic Geo-social Environment (EDGE) with over 20 years of experience in digital arts. He specializes in virtual environment building as well as researching emerging technologies. Formerly he was with the Advanced Distributed Learning Initiative ADL. As the Project Coordinator for Virtual World

Technologies, Jeff explored the facets of virtual worlds and games in support of our nation's Warfighters. Jeff's core career also includes serving as the Director of Visual Arts Development for the Program Executive Office for Simulation Training and Instrumentation (PEO STRI). He supported the Army by developing animation and multimedia programs for modeling and simulation.

Tabitha Dwyer
Tabitha Dwyer has over 22 years of experience in the simulation and gaming domains. She is currently the Creative Director and a senior developer for Cole Engineering on serious games development. She has been designing and developing the Enhanced Dynamic Geo-specific Environment (EDGE) game platform, built on both Unreal 3 and Unreal 4 game engines, with unique versions delivered to Warfighters and First Responders for enhanced training capability. She has leveraged her experience with the commercial video game industry as a publication writer to bring cutting edge technology and design ideas to simulation and training. She holds a Bachelor of Science in Computer Science from the University of Central Florida. Her prior work included traditional simulation software development on the Warfighter Simulation (WARSIM) and One Semi-Automated Forces (OneSAF) programs.

Jennie Ablanedo
Jennie Ablanedo is currently a project coordinator and research assistant at the University of Central Florida Institute for Simulation and Training (UCF IST), under contract with the Army Research Lab. She supports the Enhanced Dynamic Geo-social Environment (EDGE) game platform, built on both Unreal 3 and Unreal 4 game engines. She has over 10 years of experience in the simulation and gaming industry. She obtained her bachelor's in digital media, and her masters in modeling and simulation at the University of Central Florida with an emphasis in user experience design.
Her background and knowledge have allowed her to take an active involvement in the modeling and simulation and serious gaming communities.

Tami Griffith
Tami Griffith conducts research into the use of virtual environments for learning as a Science and Technology Manager and Team Lead for the U.S. Army Research Lab, Simulation and Training Technology Center. She is the program manager for the development of the Enhanced Dynamic Geo-social Environment (EDGE) Multiplayer Environment that replicates the tactical environment using current interactive game technology. She has twenty-five years of experience working for the U.S. Government. She has her Masters of Science in Modeling and Simulation and is working toward her PhD. Her overall goal is to improve the total user experience by leveraging emerging technology to improve Soldier learning.

Elwin Collins
Elwin Collins is the Chief of the Outreach and Exploration Branch, Federal Law Enforcement Training Centers. The Outreach and Exploration Branch is responsible for embedding within the training community, developing strong relationships with the instructor staff, and observing multiple aspects of training delivery. While embedded, the OEB staff identify gaps and challenges faced by the instructor community. Identification then leads to exploration and development of concepts, which could be in the form of methodologies or technologies. Testing, piloting and validation of the concepts are performed prior to solutions development and before any enterprise implementation action is taken. Elwin has twenty-two years of experience with the U.S. Government, a Master's of Business Administration in Project Management, and a background in electronics, security, and information technologies.

Ron Acker
Ron Acker is a Training Specialist in the Outreach and Exploration Branch, Federal Law Enforcement Training Centers. He is responsible for integrating with the training community to identify gaps in training and evaluate technology for potential implementation. Ron retired as a special agent for the Air Force Office of Special Investigations in March 2016 after spending 20 years in the Air Force where he conducted felony level criminal investigations for the Department of Defense. Ron was an instructor at the Air Force Office of Special Investigations Academy, Brunswick GA just prior to retirement. As an instructor, he provided instruction in interviewing, criminal information databases, crime scene and criminal investigation management to both newly hired and experienced agents.

A QUIZ GAME FOR JUMPMASTER TRAINING: DESIGN AND DEVELOPMENT CONSIDERATIONS

Tamara Griffith [(a)], John Surdu, PhD[(b)], Crystal Maraj, PhD[(c)]

[(a)] U.S. Army Research Laboratory Simulation & Training Technology Center
[(b)]Cole Engineering Services, Incorporated
[(c)]University of Central Florida, Institute for Simulation and Training

[(a)]tamara.s.griffith.civ@mail.mil, [(b)]buck.surdu@cesicorp.com [(c)] cmaraj@ist.ucf.edu

ABSTRACT
The Maneuver Center of Excellence (MCoE) develops doctrine and provides Jumpmaster training. They are interested in strategies to build knowledge retention and improve test performance without lowering performance standards. Students of the Jumpmaster course are required to gain proficiency in airborne operations during a short period of time. In order to meet the needs of the MCoE, the Army Research Laboratory funded the development of a tool for training called the Learning Master (LM). This paper describes a systems engineering approach to the design and development of the LM prototype. The LM system was presented to students in the Jumpmaster course to assess the usability of the application. Finally, the paper discusses potential limitations and presents the path forward to improve the LM application for future use.

Keywords: memory game, educational tool, serious game, learning master

1. INTRODUCTION

One thing more intimidating than leaping from an aircraft while in flight with only a parachute to protect you is being responsible for the safety of others making the leap. A Jumpmaster is an expert paratrooper tasked with that responsibility during airborne jump operations. Jumpmasters are responsible for teaching techniques and procedures prior to the jump. Other duties of a Jumpmaster include rigging individual equipment containers and door bundles, and inspection of all equipment, known as the Jumpmaster Pre-Inspection (JMPI) (See Figure 1) (Excellence, 2015). In short, a Jumpmaster carries the responsibility of ensuring the safety of the entire jump crew. The military Jumpmaster school prepares members from all branches of service to take on these weighty responsibilities. In Jumpmaster school, students must complete a written test during their first week before advancing into the program. Failure to pass the test results in dismissal from the course. There is a need for students to successfully complete the test in order to advance their training and to complete the course.

Figure 1: A Jumpmaster student conducts Jumpmaster pre-inspection (JMPI) to ensure the parachute is properly and safely worn.

1.1. Current Requirements and Training

Prior to admission into the Jumpmaster school, each student must meet a litany of requirements; these include the student having 12 static line parachute jumps, being in jump status for a minimum of 12 months, completing static line jumps within the past 180 days, and obtaining a recommendation from their Battalion Commander (Ft. Benning Infantry School, n.d.). Upon admission in the Jumpmaster school, each student must endure a highly demanding, three-week course that trains personnel how to ensure a safe airborne operation. During the first week of the course, students must pass a written exam in order to continue the course. The current pass rate is 70%. The rate of failure represents a loss in resources, including time, money, and manpower. Thus, the leadership who set the training standards and disseminate the training at the Maneuver Center of Excellence (MCoE), have sought strategies to increase the pass rate, without lowering the course standards. This paper describes a proposed solution to this problem in the form of a tool for training the Jumpmaster students.

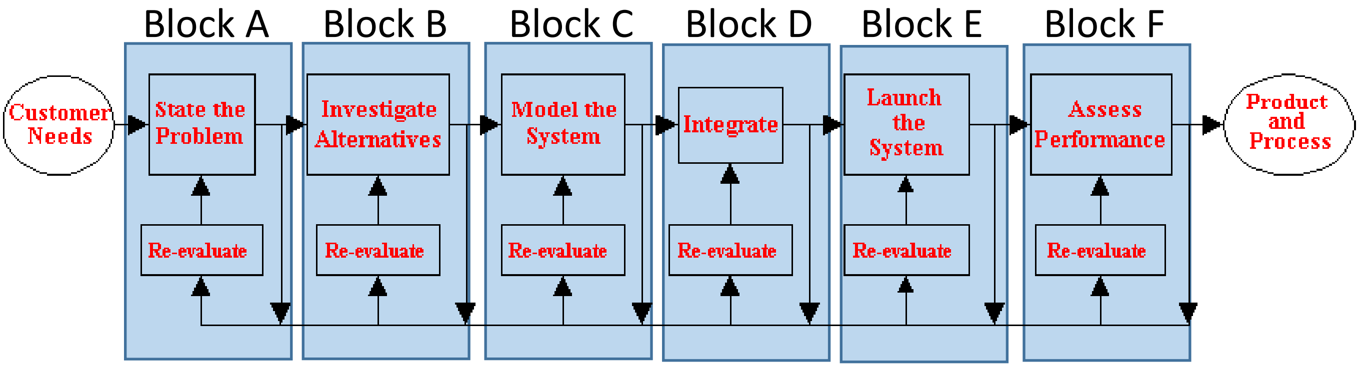

Figure 2: Systems Engineering Process. Image taken from Bahill, A.T. and Gissing, B. 1998, Re-evaluating systems engineering concepts using systems thinking, IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, Volume 28

1.2. Proposed Alternative Solution

Both cultural and technical phenomenon influenced the design of a tool to assist students in the Jumpmaster school. Firstly, the nearly ubiquitous access to personally owned, hand-held electronic devices (e.g., smart phones). Secondly, the recent emergence of online quiz games (e.g., Quizup (Quizup, n.d.)) or Trivia Crack (Etermax, n.d.)) that exercise memory while capitalizing on a student's competitive nature. From this convergence emerged the concept of a hand-held trivia game, which seeks to reinforce learning and improve the pass rate of the written exam component for the Jumpmaster course. The Army Research Laboratory (ARL), Simulation and Training Technology Center (STTC) funded a proof-of-concept of the application, resulting in a tool dubbed the Learning Master (LM).

1.3. Purpose

The purpose of this paper is to describe the design of the LM, which followed a systems engineering process for prototype development. The process involved creating a working prototype and included an initial pilot assessment.

2. DESIGN OVERVIEW

Presently, there are no formal requirements for training tools to prepare Jumpmaster students for the aforementioned written test. Due to the lack of conventional requirements, the development team paired with researcher(s) from ARL STTC to establish a set of baseline requirements for the LM. In order to advance the development of the prototype, the design team adopted the systems engineering approach by Bahill and Gissing, (1998).

After discussions with the trainers, or cadre, of the Jumpmaster school, their needs for a training tool were clarified, and the team was able to articulate the problem (Block A of Figure 2) as "develop a hand-held training tool that Jumpmaster students could use before, during, and after their Jumpmaster course to improve proficiency and first-time pass rates on Jumpmaster school tests as well as maintain proficiency throughout their careers." The trainers highlighted problem areas within the current training and this information was used for inclusion in the LM tool.

Considerations for the design goals included: a balance between pedagogy and entertainment; a need for question variety; the expected duration of sessions; learning categories; and logging capabilities, to name a few. The subsequent paragraphs present the design considerations using a systems engineering approach for the development of the LM tool. Alternative solutions were considered (Block B of Figure 2), including licensing a game engine from one of the popular developers (e.g., Quizup and Etermax) or building from scratch. In the end, it was determined that the team would license some functionality and develop the rest to provide the best mix of pedagogy and entertainment.

The team built a rapid prototype (Block C of Figure 2) in the Xamarin development environment to demonstrate the feasibility of such a handheld game. The prototype's initial design was based on the developer's personal knowledge of Army Jumpmaster training. This prototype helped informed the development team of future requirements and helped to gain buy-in by the training cadre of the basic approach.

3. DESIGN AND INTEGRATION

The first design consideration discussed was the balance between pedagogy and entertainment. It was important to design a tool that would be equally entertaining as instructional, but also promote positive training transfer (Bedwell, Pavlas, Heyne, Lazzara, & Salas, 2012) (Squire, 2003) while keeping the student engaged. Further, the design had to allow for incorporation of an existing Jumpmaster question test bank. The cadre had to be able to create, read, update, and delete questions as necessary. Although not included in the initial prototype, the designers evaluated and planned the system capabilities that would ultimately guide the students to revisit course content for missed questions using links in the application.

The second design consideration examined the need for question variety. Based on input from the trainers, the application used pre-existing test questions from their test bank. The test bank was comprised of three types of questions. The first was a standard multiple-choice question (common in commercial trivia games). The second was a multiple-choice question phrased in the form of a picture. This set of test questions was important because of its relevance to a picture-based examination, where a student identifies equipment by nomenclature. The third type of questions was true/false questions. After evaluation of the type of questions and to reduce fatigue and disinterest, the questions would be randomized for presentation, order, and responses. Future iterations of the LM will make use of Spaced Repetition (SR) based on the benefits associated with improved learning (Ausubel & Youssef, 1965; Reynolds & Glaser, 1964; Karpicke & Roediger, 2007) to optimize human performance.

Some questions were lengthy to read so it was important to provide the appropriate time interval to respond. If the time interval was too short, students would be encouraged to randomly guess. Longer times encourage deeper consideration, but may lead to boredom if students are confident in the material. At this phase, the prototype incorporated a timer that allotted sufficient time to read the longest questions and review the responses. Future re-evaluations of the prototype may benefit from implementing time as a variable based on the length of the question.

The third design consideration was the duration or length of time of the session. A snap shot of online games showed that QuizUp's sessions involve seven questions as compared to Trivia Crack's that last several rounds over several days (Russolillo, 2014). The design team evaluated each paradigm and determined that the duration should closely resemble QuizUp's seven questions, with short sessions that could fit into the student's personal time.

The fourth design consideration was learning categories. The team planned to use test questions categorized into individual practice and competitive practice sessions. Students could, then, select a certain category (to prepare for a specific test) or include all categories (where questions were chosen from each category). A competitive practice session presents identical questions to the student as well as his/her opponent. Each student receives an overall score and lists the time to respond. The student's status is recorded on the leaderboard page with rankings for total scores and total wins (See Figure 3). Future design discussion can incorporate question banks from a category that can be reused in other classes (e.g., to prepare a Soldier for the Army Best Warrior

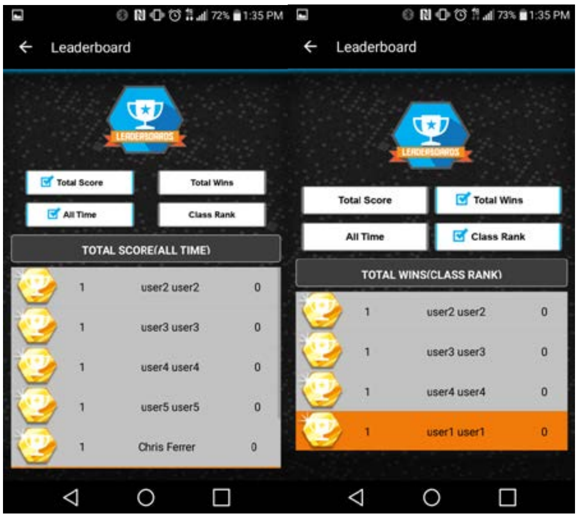

Figure 3: Left image presents total scores and right image presents total wins

competition). This creates flexibility when designing different courses within the application, allowing trainers to create new courses by combining existing courses. Ultimately, the vision of the project seeks to cross-pollinate with other courses in the MCoE.

The fifth design consideration was the logging capabilities for the LM. The design team wanted to create a back-end service that allowed for data analysis and interpretation. The data captured in the application included logging correct and incorrect responses, frequency of missed questions, response time, and the duration of time to answer. According to the trainers it was important to log student data relevant to classroom instruction for in-class feedback. This feedback provides the trainers with a summary of the student performance at any time. For instance, if a large number of students selected an incorrect response, it may be attributed to a lack of understanding of the lesson. If they, as a group, consistently select the same incorrect response, it may indicate that the question may be confusing, or that instruction was not adequate. By providing feedback, the trainers can correct misunderstandings and offer clarification. In developing the LM, the design team was able to elicit feedback from the trainers prior to making it available to students. Taking the considerations and recommendations offered, the following section details modeling the LM.

4. INTEGRATION

The Xamarin platform clients and back-end functionality were integrated and tested (Block D of Figure 2). Details of this effort are included below.

4.1. Client Software

The design of the LM includes two client software components: 1) a cross-platform application residing on the student's hand-held devices, and 2) a web-based administrative client to manage data and generate reports. The cross-platform application is built on the Xamarin platform (a commercially developed tool using a native coding language), which facilitates rapid, cross-platform development of clients for Android, iOS, Windows Phone, and Mac OS X (Visual Studio, 2013). An example of the Xamarin client application on Android and iOS is showcased.

The web-based administrative client is built using the Angular JavaScript (an open source front-end web application) framework combined with the Data-Driven Documents (D3) data visualization library. Using the administrative client, trainers are able to manage active courses, import class rosters, change question and answer values, and generate reports using current and historical student performance data.

4.2. Back-End Services

A key element of the LM tool is the inclusion of a back-end database. A back-end database allows trainers to use a web browser to directly access student performance data. For this design consideration, the LM back-end service was built on the Java platform using a Representative State Transfer (RESTful) service architecture. Under this architecture, different types of resources are exposed to the client software over a standard Hypertext Transport Protocol (HTTP) and secured using Transport Layer Security (TLS). This approach ensures users can access the LM back-end service from any type of internet connection independent of caching, proxy, Network Address Translation (NAT), and other configuration concerns. Table 1 illustrates the LM back-end services supporting the following resource objects.

5. GAINING USER FEEDBACK

In a truly agile model, the Minimum Viable Product (MVP) is presented to users frequently to gain their feedback and plan next steps. The first fully operational MVP of LM was targeted at devices spanning a range of mobile operating systems, processor speeds, and screen resolutions to assess its functionality. A list of devices are as follows:

- Apple iPhone 5S running iOS 9.3
- Nexus 5, Galaxy S4.0, LG G2 running Android 4.4 (KitKat)
- Nexus 7 tablet running Android 5.1 (Lollipop)
- Nokia Lumia 920 running Windows 10 Mobile

From the initial user feedback, it appeared that the Xamarin client functioned on all devices, although further quality assurance testing would be required prior to a wide-scale LM deployment.

Table 1: A list of resource objects with functionality

Resource	**Purpose**
Authenticated Session	Allow authentication and authorization of users for access to access-controlled back-end resources
User	An LM user authorized to use the application and/or access the administrative web client
Match	A LM application between two users with selected questions and scoring
Question	Question for use in a LM match, with allowed user response time, correct and incorrect answers, and any relevant metadata
Image	An generic image linked to by another resource
Data Report	Processed user performance data suitable for data exploration and visualization or report generation

6. LAUNCH: PROTOTYPE USER TESTING

The system was used during a Jumpmaster course at the Manuever Center of Excellence in Ft. Benning, Georgia (Block E of Figure 2). The details of the launch are described below.

6.1. Flow Model for User Testing

The Flow Model for user testing begins when a student selects the LM application then decides to start a new session, enter an existing session, challenge a friend, accept a challenge, or exit the session. If a new session is selected, then the student must select a single or multi-category option. Next, the questions are presented, followed by the student's response. Each student completes the round independently. Once the response is selected, the correct response is presented. After a round ends, points are awarded based on the number of correct responses and the length of time to respond. Higher points are awarded if answered quickly and accurately. This is repeated through seven questions. Then the student has the option to start the next round or to end the session. At this phase, the user feedback procedure is documented, however, the next phase will test the LM to assess the performance of Jumpmaster students and usability of system.

7. FEED BACK – ASSESS USER PERFORMANCE

Finally, the team used the prototype to receive user feedback and to assess the effect the game had on user performance (Block F of Figure 2). To stay within budget, the team was not able to satisfy every capability in the list of requirements. For example, questions were presented at random intervals rather than using SR spacing. This translates to some questions presented frequently with others rarely being shown. This may limit a student's full mastery of the topic or category. Another example pertains to the backend database. The database was not fully implemented in the initial prototype. This meant that the data could be made available to the developers, but the trainers would be limited in their ability to conduct queries themselves.

Another challenge stemmed from usability design. The developers didn't realize the importance of being able to zoom in to images illustrated in the LM application. Specifically, some images were overall difficult to view, in others it was difficult to determine where the arrows were pointing or what was circled. The ability to "pinch out" the image, causing it to expand (scalability), would improve the presentation and understanding of the question (See Figure 4). Understanding the encountered issues will help establish future requirements and system changes.

8. A PATH FORWARD

The focus of this paper details the steps for designing the LM in efforts to reinforce learning for students in the US Army Jumpmaster School. The LM would be used in conjunction with the current classroom-based instruction on building airborne operations skills. Emerging digital technologies and game-based training has brought forth the development of a hand-held application prototype. With input from the trainers of the Jumpmaster school, the design team was able to design using a spiral process of systems engineering to develop the LM tool.

The next phase of this research initiative requires pilot and experimental testing of the LM tool using human subjects. Specifically, the team plans to make site visits to administer the LM application and collect feedback (such as usability assessment and levels of engagement) from the Soldiers in the Jumpmaster course. Based on the provided feedback, the team plans to enhance the LM application to include a fully functional tool with logging abilities and back-end capabilities. Over time, the goal will be to expand this training tool to other branches of the Armed Forces, as well as the training community at large.

ACKNOWLEDGEMENTS

This research was sponsored by the U.S. Army Research Laboratory, Simulation and Training Technology Center (ARL STTC). However, the views, findings, and conclusions contained in this presentation are solely those of the author and should not be interpreted as representing the official policies, either expressed or implied, of ARL STTC or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government.

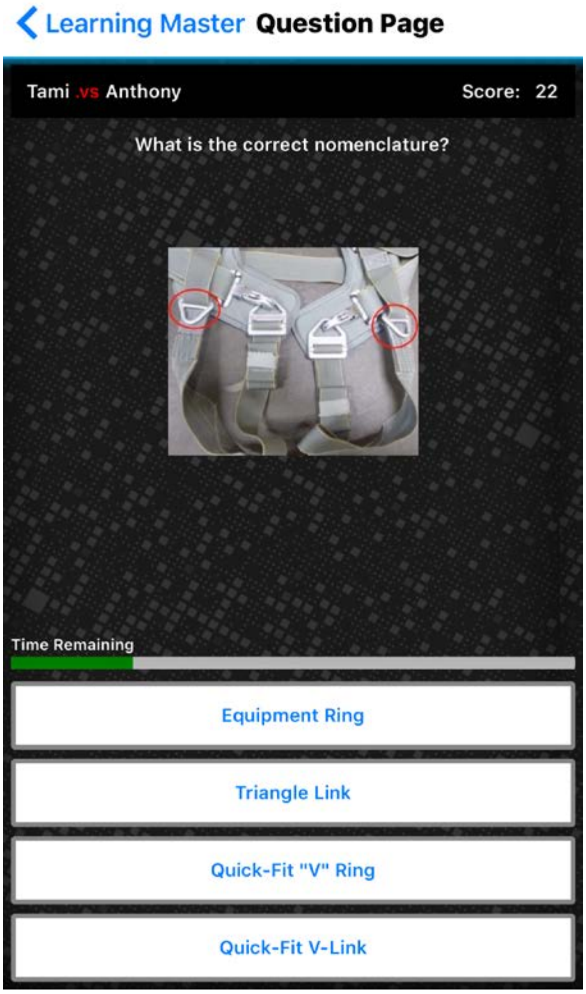

Figure 4: User design issue (lack of scalability)

The design team and researchers would like to acknowledge the support and engagement of the cadre at the US Army Jumpmaster course. Leveraging their expertise and willingness to share information has made it possible to offer a training solution to the soldiers in the Jumpmaster courses.

REFERENCES

Ausubel, D. P., & Youssef, M. (1965). The effect of spaced repitition on meaningful retuention. *The Journal of General Psychology*, 147-150.

Bahill, T. A., & Gissing, B. (1998). Re-evaluating systems engineering concepts using systems thinking. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 516-527.

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning; an empirical study. *Simul Gaming*, 729-776.

Etermax. (n.d.). *Trivia Crack*. Retrieved April 2916, from http://www.triviacrack.com/#etermax

Excellence, U. A. (2015, November 18). *Jumpmaster*. Retrieved April 25, 2016, from www.benning.army.mil/infantry/rtb/1-507th/jumpmaster/

Ft. Benning Infantry School. (n.d.). *Incoming Jumpmaster Student Information.* Retrieved from http://www.benning.army.mil/infantry/rtb/1-507th/jumpmaster/

Karpicke, J. D., & Roediger, H. L. (2007). Expanding Retrieval Practice Promotes Short-Term Retention, but Equally Spaced Retrieval Enhances Long-Term Retention. *Jornal of Experimental Psychology: Learning, Memory, and Cognition*, 33 (4), 704-719.

Quizup. (n.d.). *Quizup.com*. Retrieved November 2015, from https://www.quizup.com/en

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology*, 297.

Russolillo, S. (2014). QuizUp: The Next "It" Game App. *Wall Street Journal Live*. Retrieved November 3, 2017, from http://live.wsj.com/video/quizup-the-next-it-game-app/

Squire, K. (2003). Video games in education. *International Journal of Intelligence Simulation Gaming*, 49-62.

Visual Studio. (2013). *Xamarin 2.0 Review.* Retrieved November 12, 2017, from Visual Studio: https://www.visualstudio.com/xamarin

AUTHORS BIOGRAPHY

Tamara Griffith has 25 years of experience working for the Department of Defense and the U.S. Army. She currently conducts research into the use of game technology for learning and training as a Chief Engineer for the U.S. Army Research Laboratory Simulation & Training Technology Center. She is the program manager and technical lead for the development of the Enhanced Dynamic Geo-specific Environment (EDGE) Multi-Player Environment that leverages commercial technology to replicate tactical and first-responder environments. This effort brings the multi-billion dollar advances available from the commercial game industry to the hands of the Warfighter and First Responders for training and analysis. She has developed an architecture that encourages reuse and interoperability across various modeling and simulation tools. She has completed her Master's of Science in Modeling and Simulation and is working toward her PhD dissertation, which is focused on factors of realism in human dimension interactions in immersive environments.

John R. "Buck" Surdu currently works as a senior scientist for Cole Engineering Services, Inc. Immediately prior to joining CESI, Buck Surdu was the deputy project director within the National Security Agency as a government civilian. His last assignment before retiring from the Army was as the Chief of the Commander's Action Group for USCYBERCOM. Prior to joining USCYBERCOM, he served as the military deputy director of the Army's Communications and Electronics Research, Development and Engineering Center. Previously, he served as the Chief of Staff of the Army's Research, Development and Engineering Command and was the Science and Technology Acquisition Corps Advisor (STACA) in U.S. Forces Iraq. Before that he served as a Project Manager at the Defense Advanced Research Project Agency (DARPA) and as the Army product manager for the One Semi-Automated Forces (OneSAF) simulation system, which won numerous DoD and industry awards and leveraged cutting edge software development methodologies. He served as a research scientist at both the Army Research Laboratory and the Information Technology and Operations Center. He was also an Assistant Professor in the Department of Electrical Engineering and Computer Science at the United States Military Academy, West Point.

Crystal S. Maraj is a Research Assistant Professor employed by the Institute for Simulation and Training (IST) at the University of Central Florida (UCF). She has attained her Bachelor's degree in Psychology, as well as her M.S. and Ph.D. in Modeling and Simulation (M&S) from UCF. For the past four years, Dr. Maraj has worked for the Simulation & Training Technology Center (STTC) as a Researcher on medical simulation technology projects as well as developing and implementing empirically-based research experiments for Simulation-Based Training (SBT) platforms. Dr. Maraj has published research findings to inform the scientific and training communities to improve trainee performance and training system utility. Most recently, she is leading the Realistic Assessment of Performance in Devices (RAPID) lab that focuses on the evaluations of VR/AR technologies assessing capabilities, deficiencies, specifications, costs, maturity and risks. This information will inform the military community of the technology readiness of devices. In addition to leading the RAPID lab, Dr. Maraj also manages the Robotics Club at UCF. She mentors students as they prepare for local and national robotics competitions. Her interests in outreach programs had led her to actively participate in STEM activities that support Florida school communities.

CYBERAIMS: A TOOL FOR TEACHING ADVERSARIAL AND SYSTEMS THINKING

Erjon Zoto[a], **Stewart Kowalski**[b], **Christopher Frantz**[c], **Basel Katt**[d], and **Edgar Lopez-Rojas**[e]

[a][b][c][d][e]Norwegian University for Science and Technology, NTNU Gjøvik, Norway
[a][b][c][d][e][erjon.zoto;stewart.kowalski;christopher.frantz;basel.katt;edgar.lopez]@ntnu.no

ABSTRACT
CyberAIMs stands for Cyber Agents' Interactive Modeling and Simulation. We designed this tool in order to use it as an educational tool to teach Master students in a Cyber security course. This paper aims to describe the model and explain the design choices behind CyberAIMs in terms of associating them with the emerging concepts within cyber security curriculum, namely adversarial and systems thinking. The preliminary results indicate that the current distribution of values and entities allows most of the defense agents to avoid losing all their resources to their attack counterparts. We intend to use this tool as part of a lab with students in Information Security and further extend our target users, by including others who need training in adversarial and systems thinking. We conclude by providing rough results from running simulations with the tool and giving further directions of our future research, in order to improve the usability and level of detail for this tool.

Keywords: Agent-Based Simulation, Teaching, Cyber Security, Adversarial Thinking, Systems Thinking, Training.

1. Introduction

We can hardly pass a day without hearing of a new cyber-attack. These constant headlines of cyber-attacks have raised awareness not only among security researchers but on the whole society. Public and private organizations are calling on educational institutions to increase both the quantity and the quality of cyber security education.

To meet this call, the Joint Task Force on Cybersecurity Education (2017) has developed a new curriculum for cyber security education in high education that is both multidisciplinary and comprehensive. These new curriculum guidelines introduce some new crosscutting concepts to deal with the evolving nature of cyberspace threats. These are:

- Adversarial thinking, as a process that considers the potential actions of the opposing force working against the desired result.
- Systems thinking, as a process that considers the interplay between social and technical constraints to enable assured operations.

To address these new concepts, we developed a simulation tool, CyberAIMs, in Netlogo, Wilensky (1999). We named it in such a way to focus the attention on its main concern regarding cyberspace and the various actors interacting within that highly challenging environment. The name is an acronym for Cyber Agents' Interactive Modeling and Simulation, while it also was built to reflect certain procedures and strategies that each actor in cyberspace follows according to their own aims, as part of a higher entity or on individual basis.

We designed this tool and intend to use it further as the main component of an experiment conducted with students in Information Security, in order to further improve their adversarial and systems thinking abilities. Students participating in the experiment will be provided with a scenario and further asked to use CyberAIMs in order to give correct answers to the questions on it.

Outline of the paper We have organized the paper contents as follows. Section II will provide information on the literature that further motivated our work with the tool, while Section III will discuss the research methodology. Section IV will provide more details on the design process of the tool used and its main features. Sections V, VI and VII will conclude this paper by providing main insights from the preliminary results and relevant discussions to help the reader get familiar with the next objectives of this research process as a whole.

2. Background

In this section we explain some of the concepts that are important for the reader to understand the remain of the paper.

2.1. Modeling cyber defense and cyber war

We started designing the tool inspired mainly from a relevant work in the same area by Ben-Asher and Gonzalez (2015) and a study prepared from Ponemon Institute (2016).

Ben-Asher and Gonzalez have created a simple cyberwar game that takes place in a network of n players. Each player has two main attributes, Power and Assets. Power represents the player's cyber security infrastructure, seen also as the investment in cyber security, while Assets entail the confidential information available for use.

The Ponemon Report showed the relationships between the time spent and compensation of today's cyber attackers and the way that organizations can thwart attacks. Some relevant findings were the average cost of $1,367 on a yearly basis for the tools that an attacker needs to execute his attacks and the average time spent against different target security infrastructures, ranging from 70 up to 209 hours on average.

Ablon et al. (2014) have done an interesting work in describing the fundamental characteristics of cybercrime markets and how they have grown into their current state. They give some detailed estimates of the structure composition of both attack and defense sides in this reality.

We have used main concepts from these works and developed them further in the creation of the CyberAIMs tool, described in more details in Section III.

2.2. The socio-technical systems approach

A socio-technical system can be seen as being composed from two components: the social and the technical, as shown in Figure 1.

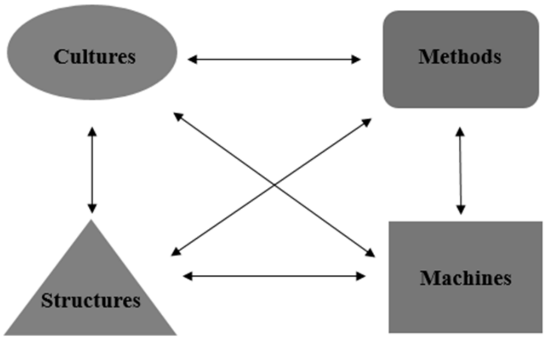

Figure 1: The typical socio-technical system, from Rogers (2000)

Each of the components can be further broken down in two subcomponents. The social component has its cultural and structural subcomponents, while the technical side has its own machines and methods as subcomponents. We have used the same approach when designing the CyberAIMs tool.

2.3. Teaching the new concepts in cyber security

Pastor et al. (2010) have done extensive research work on the available state-of-the art simulation tools that can be used on the purpose of teaching and training for cyber security. They suggest that such simulation tools should be designed to have an extremely simple user-friendly interface and, at the same time, allow the user to obtain a deep understanding of the concepts.

Adversarial thinking has already been studied as an important skill for cyber security, where Hamman et al. (2017) propose that cyber security students should learn about basic game theory concepts in order to improve their strategic reasoning abilities. Similar to Schneider (2013), our work aims to teach cyber security to undergraduate and graduate students at university level.

Systems thinking has been associated to different areas of research since several decades now, and can also be relevant for information and cyber security. There are many examples where using simulations for teaching systems thinking, such as the work from Goodwin and Franklin (1994), or the contribution from Anne Bardoel and Haslett (2004). Their seminal work motivated our work further in this paper, while aiming to use simulation as part of the curriculum developed in the field of Cyber security.

We aim to reflect the mechanisms behind the thinking processes above in the features within CyberAIMs, part of our recent interdisciplinary research work done.

3. Research methodology

3.1. Research approach

We use the design science approach to frame our research work. The chosen methodology helps create a process initiating with an artifact, which can be evaluated step after step after collecting real-time data and more feedback related to usual cyber-security events.

Design science research methodology consists of the following process steps, as in Vaishnavi and Kuechler (2004):

- The researchers gather information and build up awareness of the real world problem.
- They further suggest for a tentative design with this design as the main output.
- The researchers attempt for an artifact design, developed from the tentative design.
- The researchers evaluate the artifact with the help of performance measures.
- Finally, the design processes are completed and conclusions (results) are drawn.

The whole process iterates until the real-world situation is improved. In this research paper, we will outline the current version of the CyberAIMs tool and then provide some preliminary results we used to evaluate the simulation according to the purpose of teaching adversarial and systems thinking.

3.2. Research question

The main aim of this paper is to produce a proof of concept artifact to show if a simulation tool can affect thinking processes of a group of students in cyber security. With this artifact we hope to address the new directions suggested in developing curriculum for cyber security in education. We have devised the following research question in order to achieve the objectives mentioned above:

- Research Question (RQ): Can we design a simulation tool that helps improve adversarial and systems thinking ability on students in cyber security?

This research question helps us understand the proper features and logic behind a simulation tool that might help improve the learning process of adversarial and/or systems thinking for training and educational purposes. We believe that modeling and simulation create a good and efficient way to show results that can be further mapped to real cases of cyber events.

The modeling phase purpose is to create a normalized view of the cyber security situation, while the simulation phase allows the imitation of typical attack activities against a specific infrastructure, with specific security controls in place, grouped in sets of possible scenarios. We also believe that simulation is a method that combines theory-informed and data-driven exploration of complex interaction processes. In this work we thus propose a simulation tool, CyberAIMs, that we deem as useful in exploring the use of simulation in experiential learning environments.

4. CyberAIMs

We designed the tool in NetLogo, a programmable modeling environment for simulating natural and social phenomena. NetLogo is particularly well suited for modeling complex systems developing over time, with hundreds or thousands of "agents" all operating independently. This makes it possible to explore the connection between the micro-level behavior of individuals and the macro-level patterns that emerge from their interaction.

4.1. Agents and echelons

CyberAIMs is an agent-based simulation tool. It includes two sets of agents, namely defense and attack agents. We classified each of the groups in four distinct categories, hereinafter echelons.

Defense echelons are: Ind (individuals, ordinary people, part of a specific society, related to a specific cultural context), SMB (small and medium businesses, consisting of a relatively small number of people, with relatively low yearly income), Corp (Multinational corporations, biggest enterprises), and State (state agents, part of high-level organizations and agencies).

Attack echelons are: Kid (the script kiddies, individual hackers, part of a specific society and related to a specific cultural context), Ideol (ideological hackers, hacktivists, act on the basis on moral and ethical duty), Contract (representing cybercriminal organizations), State (state-sponsored attacker agents, representing high-level organizations and agencies, heavily engaged in nation cyberwar events during recent years).

The main purpose for defining echelons is to assign different categories on both sides, according to the level of organizational complexity that each echelon holds.

The total number of agents is variable within a range of [1,100] for both sides, reflecting contribution from Ablon et al. (2014) and additional assumptions, explained further below.

4.2. Defining agents' attributes

We started designing the tool by thinking that defense or attack actors in a potential cyber war can be represented by their own socio-technical systems. Actors will have their own *culture* - defined by certain values, traditions and laws, along with a certain *structure* - the actor's position in an organization or the whole society. They also have a certain level of access to the infrastructure already built, (*machines*), and, depending on their abilities and their will or cultural background, they can use some or other available tools, (*methods*), compared to their colleagues or potential opponents. Moreover, the type of infrastructure and tools in use should depend on the attitude of the actors or the entities they represent regarding the amount of investments made while being part of the cyber war.

Following the reasoning above, we defined three attributes that could explain the behavior and performance of the actors within our tool. The attributes were *Resources* - the budget related to cyber activities, *Skills* - level of training, literacy and awareness on cyber events, and *Motivation* - the level of self-motivation and incentives in a certain time, as in Figure 2.

Resources are most important when dealing with the technical component, spread equally between *machines* and *methods* for both attack and defense, and somewhat relevant when dealing with the *structural* subcomponent, in the process of allocating funds to different strategies applied. Figure 2 depicts this type of relationship, where all attributes

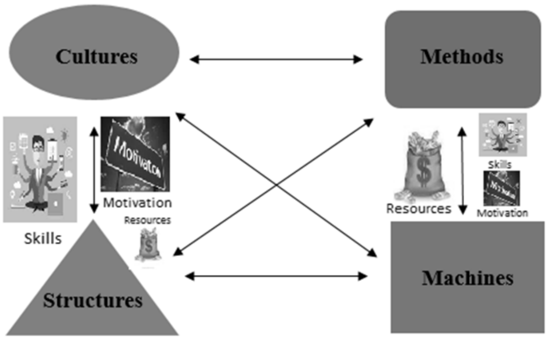

Figure 2: Attributes "produced" by the STS approach

are located and weighed according to the reasoning above.

Skills are mostly related to the social component, almost equally spread between the *cultural* and *structural* subcomponents, and somewhat relevant to the *methods* used. Motivation is generally related to the *cultural* background, but it can also be affected from the *structural* subcomponent, depending on the direct link within the different levels of management. Motivation, depending on the provided incentives, can lead to the intentional or accidental misuse of *machines*. Both skills and motivation are slightly biased towards *culture* in the social component.

4.3. Defining values for number of agents and their attributes

CyberAIMs allows the user to define initial number of agents in each side of the battlefield and also the initial value of standard deviation for each of the attributes for all agents on each side. The user can choose within a range of [1 100] on the initial number of agents on each side. He can also choose the initial standard deviation for the *Resources* units in each echelon within a range of [1 10] and for the *Motivation* and *Skills* units within a range of [1, 20] for all agents on each side.

Following the work from Ablon et al. (2014), we defined various combinations of the initial number of agents for each echelon on both sides. Thus, on the defense side, a possible combination would include 1 State, 3 Corp, 6 SMB and 90 Ind echelon agents, while on the attack side a combination would be divided between 9 State, 20 Contract, 1 Ideol and 70 Kid echelon agents. Figure 3 shows a screen-shot of the current version of the tool with a certain combination of initial agents.

In line with the research objectives, we used several estimates for the attributes defined, as shown below:

- *Resources*:
  - Defense agents
    - Ind echelon, survey data from HSB[1]
    - SMB and Corp echelon, data from Filkins and Hardy (2016)
    - State echelon, data from SIPRI Military Database[2]
  - Attack agents
    - Kid, Ideol and Contract echelon, findings from Ponemon Institute (2016)
    - State-sponsored echelon, same source as Defense State echelon agents
- *Skills*, from the Global Cybersecurity Index (GCI) Index, in Brahima (2017)

We used a heuristic model to define *Motivation* values in this version of CyberAIMs, which only included a three-levelled scale from Low to High, as shown in Table 1. In the next version of the tool, we intend to add a number of different motivation theory models, explained further in the last section.

We performed several types of computations to put all values of the attributes within the predefined range. Hence, we multiplied the source data on *Skills* by 100, while mapping the source data on *Resources* within the [1 100] range required more transformations. As an example, an individual spending $1000 in the real world has 25 units of *Resources* within the tool, while a state spending $1 billion would have 75 units of *Resources*. On the other hand, agents representing Singapore, the country with the highest GCI score, would have on average 92 units of *Skills*.

Since echelons should be divided according to organizational complexity, this could be linked to their level of *Resources*. Hence, we defined four sets of *Resources* values on the agents of each side. Meanwhile, as the GCI Index divides countries in three separate categories according to their overall performance, we used similar reasoning for creating three levels of the *Skills* parameter, namely Low, Medium and High. A similar classification was used for *Motivation* values.

4.4. Main features and formulas

The tool runs in several steps, called ticks. The ticks usually represent a fixed period, i.e. hours, days, months etc. We chose the period of 3 days as the time equivalent of one single tick. The reasoning behind this relates to the data from Ponemon Institute (2016), where the minimum time required to perform a successful attack was almost 3 days (70 hours) on average. We set a maximum value of ticks equal to 120, so as to simulate interactions within a short-term period of up to 1 year.

[1] https://www.munichre.com/HSB/2017-cyber-survey/index.html

[2] SIPRI Military Expenditure Database 2017, https://www.sipri.org/databases/milex

ISBN 978-88-85741-12-6; Bruzzone and Sottilare Eds.

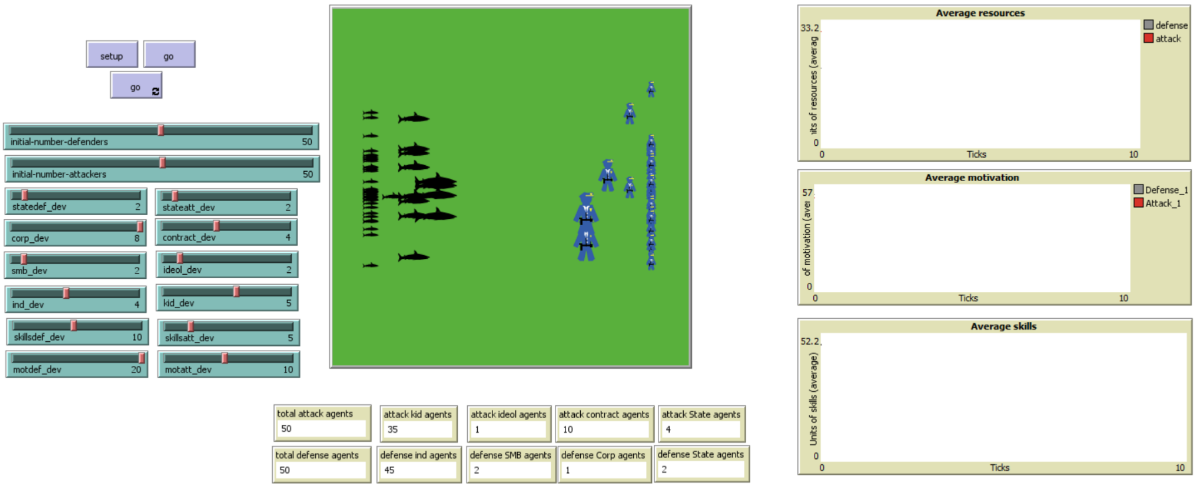

Figure 3: Screenshot - CyberAIMs

Table 1: Distribution of attributes' values

Attribute	**Side**	**Echelon/ Level**	**Range of values**
Resources	Defense	Ind	1-31
		SMB	1-40
		Corp	40-70
		State	60-100
	Attack	Kid	1-37
		Ideol	15-37
		Contract	35-67
		State	60-100
Skills	Attack/ Defense	Low	1-30
		Medium	31-70
		High	71-93
Motivation	Attack/ Defense	Low	1-33
		Medium	34-66
		High	67-99

Thus, the Ponemon Institute Report defines that whenever the defense side has ”typical”, average, *Resources*, the attackers could execute an attack within 3 days on average, which, depending on other circumstances (attributes), could be shortened or extended. Our rationale defines that the average time could change due to relative *Skills* and *Motivation* between the two sides in action. This is also true when dealing with the defense actors having ”excellent” infrastructures, meaning high level of *Resources*.

After each tick, the tool updates *Resources* on both sides only if there is a successful attack undergoing, which depends on the value of the relative power between opponent sides' agents, calculated by the formula below:

$$RelPower = LDR_{DA} + LDS_{DA} + LDM_{DA} \quad (1)$$

In equation 1, *LDR* defines the difference between *Resources* levels, *LDS* is the difference between *Skills* Levels, *LDM* the difference between *Motivation* Levels, *D* stands for Defense and *A* for Attack side agents.

In this case, we assume that the attack agents have enough initial information on their opponents' *Resources* values, while *Skills* are harder to disclose as long as there is no attack. The relative strength of defense or attack agents as measured by equation 1 reflects a visible mechanism for promoting adversarial thinking, while the use of all attributes in the same weight means there should be no initial predefined biased effects on using one attribute over the other when analyzing further the data from the simulation runs. This supports the whole view of the process, as related to systems thinking.

Thus, the attack agents can only affect the outcome by changing their attitude, expressed in *Motivation* units, in order to defeat the target within 3 ticks. Depending on the time frame, *Impact* values of the first attack should range from 0,206 for the slow defeat to the value of 1 for the quick defeat. Attack agents should increase *Motivation* by one level each tick if the *Impact* value of their first potential attack is lower than 0,206, otherwise they can not defeat their opponent within 3 ticks and would potentially have to quit. This would be the main condition towards performing the attack or not. Impact values depend on the difference between Levels of attributes and are highly related to the *RelPower* value described above.

The tool updates after each attack the *Resources* values of the defense agents by subtracting *Impact*Resources* units from it and subsequently updates respective value of the attack agents by adding the same units to them, as shown in Figure 4. It also updates agents *Skills* and *Motivation* values depending on the *Impact* of the attack.

Reflecting on the work from Pastor et al. (2010),

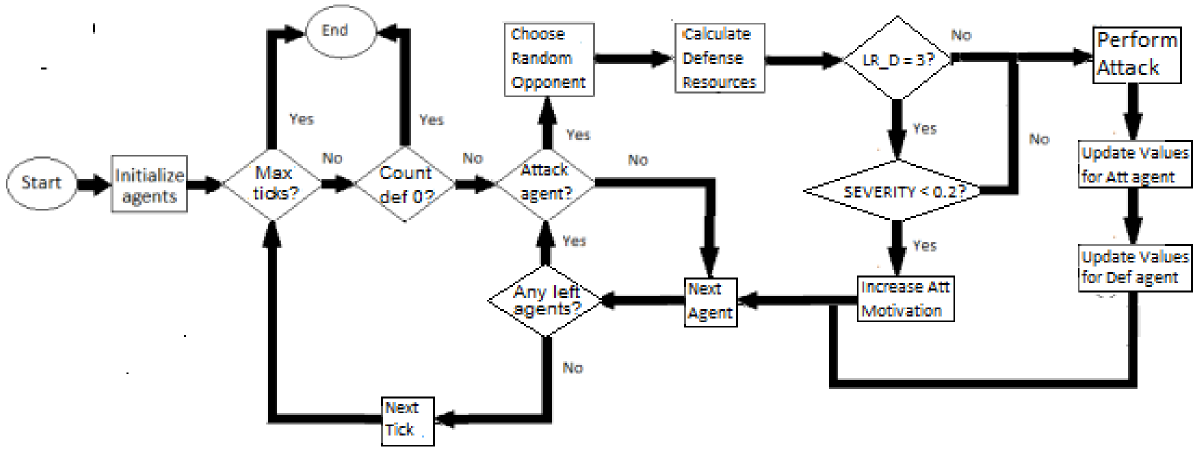

Figure 4: Flowchart explaining CyberAIMs cycles

the CyberAIMs tool can be considered under the category of Canned Attack/Defend Scenarios, as it is still standalone and can be used also as a game. It is still not remotely usable and not yet virtualized, though we already consider these as real objectives for the future releases of the tool.

The tool's level of scalability is yet to be tested, while his intended current audience are the students in Cyber security courses, with future targets including other stakeholders, especially in the critical infrastructure areas. The teaching objectives are related to the processes of adversarial and systems thinking, while the learning curve will be tested further. Finally, we intend to use surveys and scenarios to assess the level of usability and details that the CyberAIMs tool shows.

5. Results

We have conducted a systematic variation of the main attributes included in our tool, along with the initial number of agents on both sides. The values were randomized for the *Resources* attribute within each Echelon on both sides and also for *Skills* and *Motivation* attributes in their full range, according to the distribution of values defined in Table 1.

For computational reasons, we used for the current version a limited amount of single data points for the standard deviation values of all attributes and for the initial number of agents for each side, as shown in Table 2. This approach provided around 1 million rows of different combinations.

Preliminary results indicate that there are cases of defense agents going offline, but, overall, the majority of defense agents is able to avoid complete defeat in the cyberspace environment. This means that all runs end in 120 ticks.

Rough results show that in only 36% of the cases, some of the defense agents cannot "survive" until the final step. From these cases, in most cases the agents going offline were part of the Ind echelon (93%), followed by the SMB echelon (7%). Other echelons were stable in all runs.

An interesting finding relates all cases of SMB echelon agents not "surviving" with higher values of standard deviation on their *Resources* values. These cases were also related to ranges of *Skills* and *Motivation* values that were further from the predefined mean averages, so extreme values on these attributes on the Defense side reduce the chances of respective agents avoiding successful attacks, while a higher variation of values on the attack side would further increase the former agents' vulnerability.

Table 2: Selected data points of input variables

Attribute	Side	Echelon	Values
Total agents	Attack Defense	All echelons	25, 50, 100
Resources	Defense	Ind	Mean: 16 SD: 4, 8
		SMB	Mean: 20 SD: 5, 10
		Corp	Mean: 55 SD: 4, 8
		State	Mean: 80 SD: 5, 10
	Attack	Kid	Mean: 19 SD: 4, 8
		Ideol	Mean: 26 SD: 4, 8
		Contract	Mean: 51 SD: 4, 8
		State	Mean: 80 SD: 5, 10
Skills	Attack Defense	All echelons	Mean: 47 SD: 5, 10, 20
Motivation	Attack Defense	All echelons	Mean: 50 SD: 5, 10, 20

In most cases for the Ind echelon, the changes in final number of defense agents were related to variations of the *Resources* attribute on their side. This can be related to their low *Resources* values as compared to the whole range of echelons in the model. The closer their *Resources* are to the average values, the impact from other attributes becomes more visible.

6. Conclusions

This paper aims to contribute on recent research done in respect to the learning benefits of simulation tools in cyber security education. We introduced CyberAIMs, an agent-based simulation tool designed in NetLogo. Preliminary results of this proof-of-concept paper indicate that the current tool allows more than 90% of the defense agents to avoid losing all their resources to their opponents, while the ones not reaching the final step are still able to avoid this situation in more than 60% of the cases.

According to the distribution of values and the logic behind each execution cycle, agents from Ind and SMB echelons are the most vulnerable to opponent attacks, thus their low values of *Resources* seem to contribute to this result. However, this impact grows stronger whenever their values of *Skills* and especially *Motivation* stand further away from the predefined mean values, i.e. when the standard deviation has a value of 10 or larger.

Our results indicate that *Resources* are important for "surviving" in cyberspace, but having *Skills* and *Motivation* values close to average or higher, in order to oppose successfully their opponents' relative power, could help improve the performance of the defense side.

Thus, CyberAIMs is designed to have a positive effect on the students understanding of systems and adversarial thinking. Anyway, we aim to further evaluate this statement by introducing this tool to various groups or students, as well as other types of users in the future, as part of an iterative process that stems from the research approach we are following.

7. Future work

We intend to use CyberAIMs as part of various simulation-based experiments with different samples of students in Information Security. We have already created a scenario[3] based on a real cyber case occurred recently, where an Iranian state-sponsored group was successful in targeting critical infrastructure entities in the US, Saudi Arabia and South Korea. We will link the scenario with pre- and post-simulation surveys to validate further the level of usability and details that the CyberAIMs tool shows, as related to adversarial and systems thinking.

[3] https://thehackernews.com/2017/09/apt33-iranian-hackers.html

In the next releases of CyberAIMs, we intend to look deeper into the *Motivation* attribute, by conducting a more detailed literature review on the theories explaining attack actors motivation, such as the ones related to the MOMMs taxonomy in Bologna (1981), and other theories explaining defence actors motivation, such as the protection theory from Rogers (1975).

Another direction will aim to incorporate the work from Norman and Koehler (2017), trying to model the high-level entities by using the concept of a complex adaptive system, made up of interacting components (agents) that adapt their behavior over time in reaction to environment changes and to each other. We will also address the issues related to cyber security investments, as explained in Rowe et al. (2013), as a means to define how can *Resources* and *Skills* levels change in a mid-term and long-term period.

There are other factors that need to be further incorporated in the model, such as getting the defense agents fully or partially recover from defeat or defining when they are able to detect an ongoing or past attack and how can they respond to it. The next version of the tool will include a list of several strategies that individuals and other entities follow in their cyber defense policies, using contributions from several works, such as Filkins and Hardy (2016) and AlSabbagh and Kowalski (2017), in order to define the best strategy or combination of strategies that allows defense agents to overcome potential cyber attacks from agents of different echelons with different attack strategies in a multidimensional environment.

Acknowledgments

This work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme, in the Information Security and Privacy Management research group, and supported further from the Systems Security and Digital forensics research groups, at the Information Security and Communication Technology Department, NTNU Gjøvik, Norway. We would like to express our gratitude to the other colleagues from the Department supporting this research.

References

Lillian Ablon, Martin C Libicki, and Andrea A Golay. *Markets for cybercrime tools and stolen data: Hackers' bazaar*. Rand Corporation, 2014.

Bilal AlSabbagh and Stewart Kowalski. Socio-technical siem (st-siem): Towards bridging the gap in security incident response. *International Journal of Systems and Society (IJSS)*, 4(2):8–21, 2017.

E Anne Bardoel and Tim Haslett. Success to the successful: The use of systems thinking tools in teaching ob. *Organization Management Journal*, 1(2):112–124, 2004.

Noam Ben-Asher and Cleotilde Gonzalez. Cyberwar game: A paradigm for understanding new challenges of cyber war. In *Cyber Warfare*, pages 207–220. Springer, 2015.

J Bologna. Momm's (motivations, opportunities, methods, means)-a taxonomy for computer related employee theft. *Assets Protection*, 6(3):33–36, 1981.

S Brahima. Global cybersecurity index 2017. *International Telecommunication Union (ITU)*, pages 1–77, 2017.

Barbara Filkins and GM Hardy. It security spending trends. *A SANS Survey. SANS Institute*, 2016.

Jack S Goodwin and Stephen G Franklin. The beer distribution game: using simulation to teach systems thinking. *Journal of Management Development*, 13(8):7–15, 1994.

Seth T Hamman, Kenneth M Hopkinson, Ruth L Markham, Andrew M Chaplik, and Gabrielle E Metzler. Teaching game theory to improve adversarial thinking in cybersecurity students. *IEEE Transactions on Education*, 60(3):205–211, 2017.

Joint Task Force on Cybersecurity Education. Cybersecurity curricula 2017 - curriculum guidelines for post-secondary degree programs in cybersecurity - csec2017 v. 0.95 draft. Technical report, November 2017.

M. D. Norman and M. T. K. Koehler. Cyber Defense as a Complex Adaptive System: A model-based approach to strategic policy design. *ArXiv e-prints*, June 2017.

Vicente Pastor, Gabriel Díaz, and Manuel Castro. State-of-the-art simulation systems for information security education, training and awareness. In *Education Engineering (EDUCON), 2010 IEEE*, pages 1907–1916. IEEE, 2010.

Ponemon Institute. Flipping the economics of attacks. Technical report, January 2016.

Marc Rogers. A new hacker taxonomy. *University of Manitoba*, 2000.

Ronald W Rogers. A protection motivation theory of fear appeals and attitude change1. *The journal of psychology*, 91(1):93–114, 1975.

Brent R Rowe, Igor D Pokryshevskiy, Albert N Link, and Douglas S Reeves. Economic analysis of an inadequate cyber security technical infrastructure. *National Institute of Standards and Technology Planning Report*, pages 13–1, 2013.

Fred B Schneider. Cybersecurity education in universities. *IEEE Security & Privacy*, 11(4):3–4, 2013.

Vijay Vaishnavi and William Kuechler. Design research in information systems. 2004.

Uri Wilensky. Netlogo. evanston, il: Center for connected learning and computer-based modeling, northwestern university, 1999.

AUTHORS BIOGRAPHY

Dr. Erjon Zoto

Erjon Zoto is currently a post doctoral fellow researcher under the ERCIM 'Alain Bensoussan' Fellowship Programme, hosted by the Information Security and Privacy Management group at NTNU in Gjøvik, Norway. Erjon obtained his PhD in Information Systems at the University of Tirana, Albania, and his current research area is information security economics and related interest topics: Agent-Based Simulation and Modelling, Systems thinking and Adversarial Thinking. Erjon has a Bachelor degree in Economic Informatics from University of Tirana and a joint Masters degree in European Economic Studies from University of Tirana and Bamberg University in Germany.

Professor Stewart Kowalski

Stewart Kowalski has over 35 years of industry and academic experience in information security and has worked for a number of large international companies including Ericsson, Telia Research, Huawei, Digital and HP. He has also taught and researched information security at a number of universities, including the Royal Institute of Technology (KTH), Stockholm School of Economics, and Stockholm University Department of Computer and Systems Sciences. He is currently Professor of Information Security at the Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Norway. He is also the current head of the Information Security Management Group at the Center for Cyber and Information Security, CCIS, www.ccis.no and Program Director for 4 Master Programs.

Assoc. Prof. Christopher Frantz

Christopher Frantz is an associate professor at the Department of Computer Science at the Norwegian University of Science and Technology (NTNU). He has an industrial background in the area of software engineering and holds a PhD in the area of agent-based modelling from the University of Otago, New Zealand. His expertise lies in the area of agent-based modelling and simulation, with specific focus on normative multi-agent systems and institutional analysis. A parallel research theme is the model-driven development of institutions using blockchain technology.

Assoc. Prof. Basel Katt
Basel Katt is an Associate Professor at the Norwegian University of Science and Technology (NTNU), the Department of Information Security and Communication Technology. Basel Katt holds a Bachelor degree in Electronics and Communication Engineering, and received his M.Sc. degree in Computer Science from the University of Essen-Duisburg, Germany, and his PhD degree in Computer Science from the University of Innsbruck, Austria. His research interest lies in the areas of security engineering, security testing, model driven security, security education, access control, and anomaly detection. In 2015 He joined NTNU in Gjøvik and he is leading the System Security research lab.

Dr. Edgar A. Lopez-Rojas
Edgar Lopez-Rojas is currently a post doctoral researcher at the Digital Forensics group at NTNU in Gjøvik, Norway. Edgar obtained his PhD in Computer Science at Blekinge Institute of Technology and his research area is fraud detection and related interest topics: Multi-Agent Based Simulation, Machine Learning techniques with applied Visualisation in the domain of retail stores, mobile payments and Anti Money Laundering (AML) for financial transactions. Edgar has a Bachelors degree in Computer Science from EAFIT University in Colombia and a Masters degree in Computer Science from Linköping University in Sweden.

MODEL AND SIMULATION OF A REAL WATER DISTRIBUTION NETWORK TO SUPPORT EMERGENCY PLANS

Luisa Lavalle[(a)], Tatiana Patriarca[(b)], Bernard Daulne[(c)], Olivier Hautier[(d)], Ester Ciancamerla[(e)]

[(a),(b),(e)]ENEA – Agenzia Nazionale per le nuove tecnologie, l'Energia e lo sviluppo economico sostenibile
[(c),(d)]SWDE – Société Wallonne Des Eaux

[(a)]luisa.lavalle@enea.it, [(b)]tatiana.patriarca@enea.it, [(d)]ester.ciancamerla@enea.it
[(c)]Bernard.Daulne@swde.be, [(d)]Olivier.Hautier@swde.be

ABSTRACT
This paper presents a water distribution network model to evaluate the impact of adverse events, such as faults and/or cyber-attacks, on a real water distribution system in a wider context which involves the interdependency with the electrical grid, in the frame of the Horizon 2020 project ATENA. The model has been developed by using a commercial simulator, which can address both the electrical and the water domain. Specific features and modules have been added to the simulator, in order to calculate the water level in tanks – an important and missing metric to support emergency plans. The interdependency among the electricity grid and the water network is considered throughout pumps, which are at the same time electrical loads and hydraulic devices. Two use cases, dealing with faults or cyber-attacks against the electrical grid affecting critical pumps or treatment stations, are investigated and the simulation results are reported.

Keywords: water distribution systems, domain simulators, modeling and simulation, critical infrastructures

1. INTRODUCTION

The protection of Critical Infrastructures (CIs) has been considered a paramount objective by all modern countries around the world for many years. Identifying weaknesses of a CI in order to put in place effective prevention or mitigation strategies is one of the most critical challenges, made even bigger by the fact that all the infrastructures are interconnected and rely on each other to deliver their service, therefore a new set of hidden weaknesses is related to the lack of the services they are dependent on. Although we rarely experience severe service outages and feel that the infrastructures are robust enough, the reliability we are used to is jeopardised by both the reduced dependability of the interconnected infrastructures (for example, the electrical infrastructure is challenged by the integration of an increasing share of renewable distributed generation due to economic and politic requests) and the increasing cyber threats due to the socio-political changes.

This wider framework is taken into account inside the European Union's Horizon 2020 research and innovation project ATENA (ATENA Horizon 2020 project, 2016-2019), where three interdependent CIs – an active electrical grid, a gas network and a water distribution network – are investigated in order to develop a suite of tools able to early detect and react to cyber-physical anomalies (malicious attacks or unexpected faults) on the Information and Communication Technology (ICT) devices and/or on the SCADA (Supervisory Control And Data Acquisition) of the interdependent CIs. The work, started within EU FP7 CockpitCI project (CockpitCI 2012), refers to an advanced scenario taking into account the integration of small distributed generation sources, storage devices, passive customers and aims at improving the awareness of the network operators, interested in providing real time optimization of network flows while satisfying their evolved customers.

To achieve the ATENA objective, a big effort in modeling and simulation is undergoing to represent the CIs behaviour in normal and critical operation, taking into account the physical devices, the SCADA functionalities and processes as well as the CIs interdependencies at physical, geographical, cyber and organizational layers. In fact, models will provide knowledge and algorithms to feed the ATENA suite of tools. For such an understanding, realistic use cases and relevant performance indicators have to be investigated.

To deal with multiple infrastructures, Pederson, Dudenhoeffer, Hartley and Permann (2006) distinguish two simulation approaches, Integrated Model and Coupled Model. In the Integrated Model approach, multiple (interdependent) infrastructures (electricity, telco, water, gas, transport etc) are modeled in a single framework, while in the Coupled Model approach multiple simulators, each one modelling a single infrastructure, are connected and synchronized. Despite being more challenging due to the difficulty of interfacing and synchronizing different simulators, the latter approach allows to integrate domain specific high

fidelity models for the single infrastructures, giving more accurate results. In this paper, we refer to this approach and report the modelling and simulation activity carried out in the drinking water domain.
The safety-critical role played by water distribution systems makes them attractive targets for terrorism and cyber threats (Horta 2007; Dakin, Newman and Groves 2009) but, as Taormina, Galelli, Tippenhauer, Salomons and Ostfeld (2017) found that the same hydraulic response can be obtained by completely different faults, the chosen policy in this paper is to investigate the impact of the different disruptions, independently by the specific occurred cause and, despite a number of studies employing a mathematical approach (for example, Perelman and Amin 2014) to assess the vulnerabilities, it has been chosen to develop a simulator to follow the evolution in time of the faults and their impact on the interdependent CIs. This approach has been successfully followed by some research groups in US (Leclair, Pasqualini, Dreicer, Toole, Urban, Bent, Mcpherson and Hengartner 2014; McPherson and Burian 2005), who pointed out the importance of having credible and traceable data in order to provide a trustworthy assessment.
The work presented in this paper is carried out within the ATENA project. The authors think this work is at the forefront of the modelling and simulation research because of both the size of the modelled network and the quality of the acquired data as well as the detailed model and the novelty of the approach to assess potential weaknesses of a water distribution network. In fact, to the authors' knowledge, very few studies exist analysing large water distribution networks and they are quite difficult to compare because they focus on different aspects and use different technique and tools to solve the hydraulic equations. For example, Boano, Scibetta, Ridolfi and Giustolisi (2015) analysed a real large water distribution network (approximately 3000 pipes longer than 170 km in total and 50000 users) but they focus on pump operation optimization and employ a proprietary tool – WDNetXL – basically developed in Matlab The-MathWork and interfaced with MS-Excel. As another example, Ramesh, Santhosh and Jagadeesh (2012) consider a total area of 165 km^2 but model the network with few nodes (about 110).
The paper is organized as follows. Section 2 describes the real water distribution network to be modeled. Section 3 details the main points of the elicited model and the implemented simulator of the water distribution network. Two use cases are briefly described in Section 4 while simulation results of the network in normal operation and under the fault conditions identified in the related use cases are discussed in Section 5. Conclusions and envisaged future work are drawn in Section 6.

2. THE WATER DISTRIBUTION NETWORK

The portion of the water distribution network has been proposed by Société Wallonne Des Eaux (SWDE), partner of the ATENA project and Belgium Water Network operator, and it is a simplified version of the Belgium real water network, covering the water production from the Eupen and La Gileppe dams, the transportation through adduction pipelines and the distribution to residential and industrial sub-networks. Figure 1 shows a "Z-view" representation of the water network under study whose purpose is to illustrate the altitude of the assets and their position into the network in order to describe the process while the geographical view of the original network is shown in Figure 2.

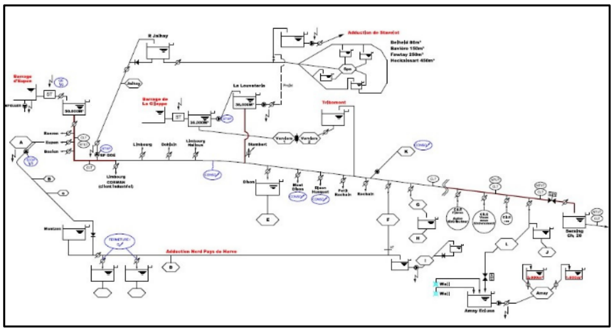

Figure 1: Water Distribution Network Z-view

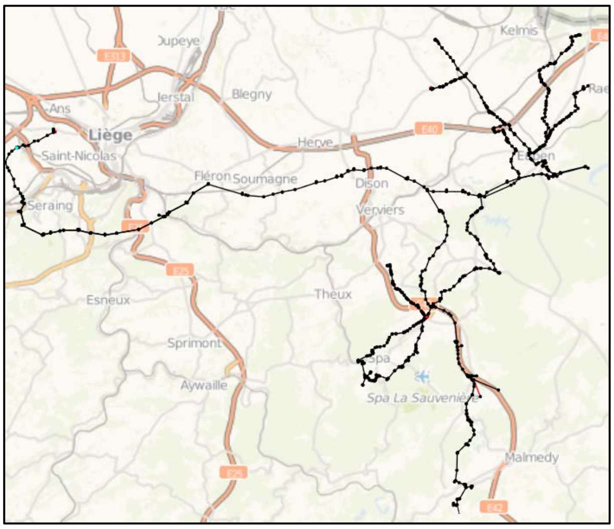

Figure 2: Water Distribution Network Geographical View

The modeled water network (shown in Figure 2) covers an area of more than 1600 km^2, it has nearly 1000 nodes and pipes length more than 200 km. This topology, featuring meshes as well as multiple sources and being extensive enough, allows to identify multiple failure scenarios (detailed in Section 4) and different reconfiguration options. One of the modeling activity's purposes is to support the operator in mitigating the effect of adverse events. Among the features, it is requested that, in case of service disruption due to either natural or cyber cause, the model will be able to compute the remaining storage capacity.
The network is supplied by 16 water tanks (deployed along the network) and 2 reservoirs. Most of the tanks (highlighted by red circles in Figure 3) are connected in a "chain", meaning that one tank supplies not only its

own consumers but also one or more tanks downstream, which, in fact, are (major) consumers too.

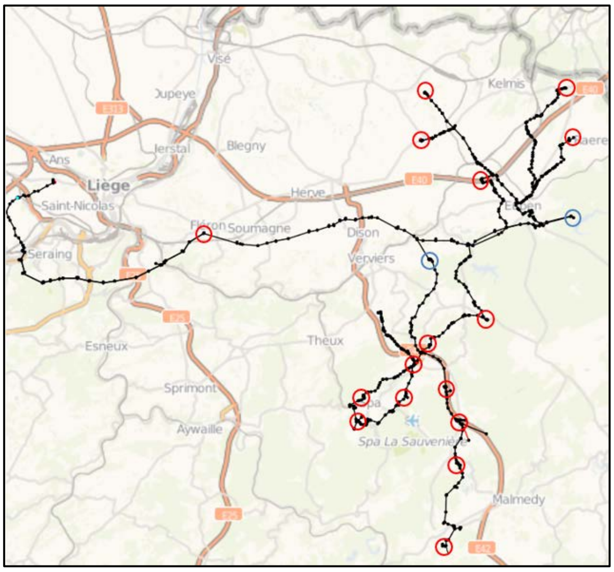

Figure 3: Water Tanks and Reservoirs in the Network

At each reservoir (highlighted by blue circles), water is filtered and chemically processed by automated treatment stations to meet the drinkable quality requirements. Generally, the water is then transported by gravity through the pipeline and, when geographical constrains require it, extra pumps - electricity loads – are installed to lift water from lower altitude to storage asset located on higher spot. As a general rule, the purpose of a water distribution system is to fulfill consumers demand while ensuring appropriate quality of the delivered water. In our scenario the water network provides drinking water to more than 100 “aggregated consumers” with a variety of profiles. It’s worth noting that from now on a consumer is a node which simulate a set of real consumers (perhaps 10, 100, 1000,.. real consumers), and have associated a base demand. The consumer flow depends on time of the day, day of the week, etc and is modelled by flow profiles.

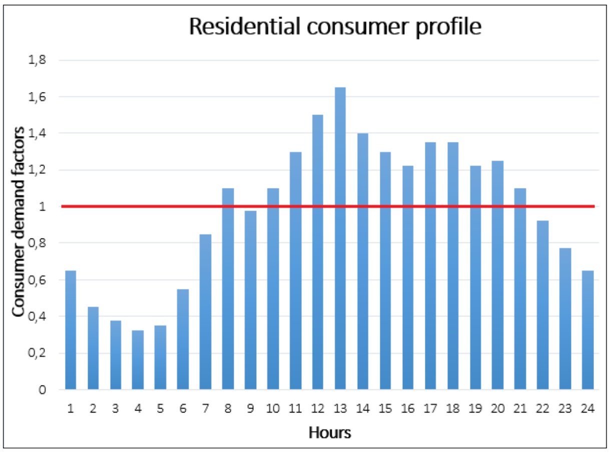

Figure 4: An example of Residential Consumer Profile

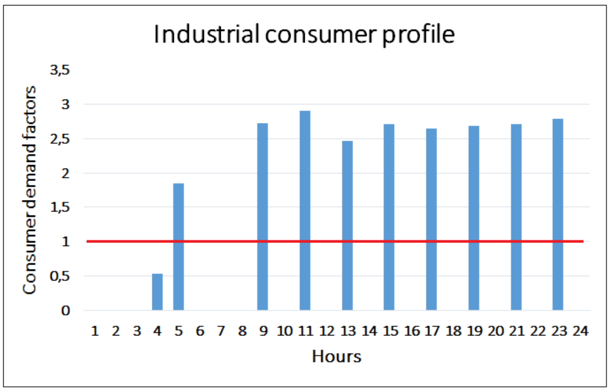

Figure 5: An example of Industrial consumer profile

As an example, water consumer flow in the 24 hours of a day for residential and industrial consumers is shown in Figure 4 and Figure 5 respectively.

3. WATER DISTRIBUTION NETWORK MODEL

SWDE provided the Consortium with the topology and physical data of a large subnetwork under their operation as well as information about the support the simulator should provide in order to be useful. As a general rule, the purpose of a water distribution system is to fulfil consumers demand while ensuring appropriate quality of the delivered water.

3.1. Employing a commercial domain simulator

The model of the real water network has been implemented with the commercial domain specific simulator PSS-Sincal by Siemens (PSS-Sincal). The reasons for this choice are two-fold. According to our experience in past projects (Balducelli, Di Pietro, Lavalle and Vicoli 2007), using a commercial simulator makes a lot easier to elicit the knowledge and existing schemas in the control room as well as sharing or exporting the results because the graphical interface and the needed parameters are usually familiar. In fact, the level of abstraction of a commercial simulator is closer to the infrastructure and then closer to the infrastructure operator’s background, therefore he can understand and use its results more easily than the results of other modeling techniques based on heavy math aspects. Furthermore, we use PSS-Sincal instead of a widely-known open source water domain simulator like EPANET (Rossman 2000) in order to make easier the integration among different domain simulators. In fact, PSS-Sincal provides specific dedicated modules for the electrical, water and gas domains which are in the scope of the ATENA project.

Running a commercial simulator requires that all the network elements of the water systems have all their own parameters properly assigned. More in details, the supply sources are modeled as water tanks, each one with their own volume and maximum water level, as well as their own node characterized by geographical coordinates and elevation (above the sea level). Pipes require length, diameter and roughness as parameters, and are connected through nodes, which are geographically localized. In PSS-Sincal, pumps – which are mandatory when altitude difference cannot be compensated by hydraulic head – can be centrifugal or reciprocating and in both cases require the characteristic curve. Valves are another important network element. They can be either sliding or non-return or flow control or pressure control valves depending on their role in the network. Finally, consumers are linked to nodes and are described by base demand and profiles. The base demand is the average water flow requested (an average of 100-120 m^3/day/person), while profiles take into account the hourly different demand. Figure 6 shows a zoom on a portion of the modeled network featuring the network elements described above.

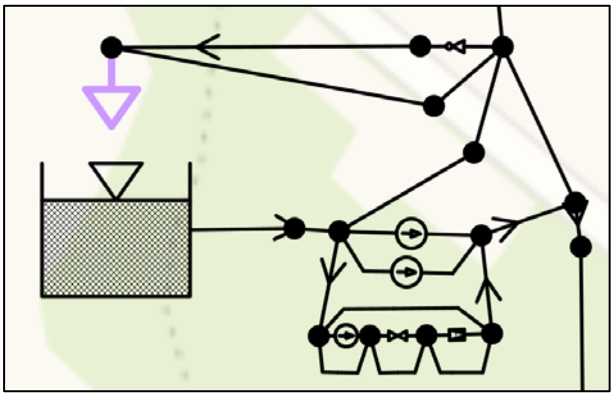

Figure 6: Snapshot showing the Network Elements

3.2. Adding new features and modules

As the purpose of a water distribution system is to fulfil consumers demand while ensuring appropriate quality of the delivered water, the operator is mostly interested in knowing how many hours are needed until one or more water tanks will be empty, given an initial configuration of the water distribution network. To capture the way in which the tanks are filled, PSS-Sincal has a specific module called "Water Filling" which allows to describe how tanks are filled (start/stop level thresholds and pump characteristics), but it doesn't take into account where water is adducted from nor if there is water in the adduction pipe. Unfortunately, due to the requirement to evaluate tank's emptying time, an estimation of the supplied water at each tank as accurate as possible is mandatory and the values calculated by the "Water filling" module were not accurate enough for our scope. Moreover, the "Water filling" commercial module stops the simulation as soon as one of the tanks is empty and this was not the best option for our purpose as we wanted to follow the complete evolution.

In order to overcome these restrictions we decided to refine the model, to interface PSS-Sincal with an ad-hoc developed software in order to be able to use its features and to develop additional modules to work around the standard algorithms.

More specifically, because PSS-Sincal doesn't model the water filling the tank as a network consumer and raises the water level in the tank as if water came always from an infinite source therefore the calculated amount of water which is taken from the tank in a certain amount of time (water flow Qsupply [l/s or m^3/h]) wouldn't be accurate enough to calculate the remaining storage capacity.

To overcome that we added a fictitious «consumer», Q_{in}, taking into account the water needed to fill the tank. Figure 7 shows an example of such an insertion.Figure 7: Model of a Water Tank with Fictitious Consumer (in lilac)

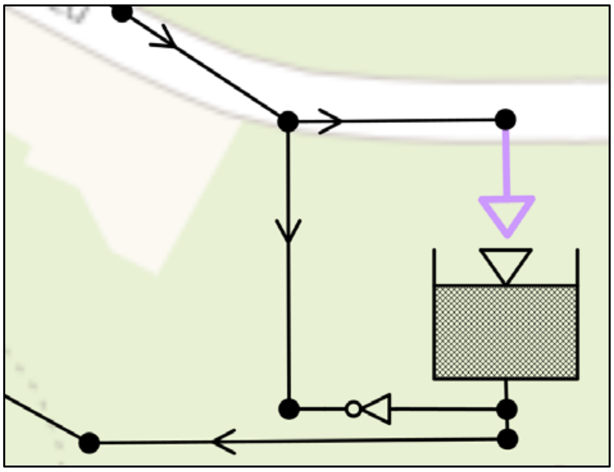

Figure 7: Model of a Water Tank with Fictitious Consumer (in lilac)

For each tank, at each time-step of the simulation this fictitious consumer models the amount of water flow (Q) requested by the "filling pumps" when the level of the water in the tank undergoes specific thresholds.

The values of Q_{in} are forced by a software module developed on purpose in order to control and move forward the simulation.

The calculation of the correct values take into account the supplied ordinary consumers profile which change at every hour.

Another aspects to be taken into account when implementing the approach described above is that the fixed amount of water provided can be "too much" when the water tank is almost full. The algorithm would force a value which would make the tank to overflow, while in the real world float valves would stop the pump. The longer the simulation time-step, the bigger the error. Having in mind that the most common simulation time-step is 1 hour – therefore the error could have been relevant – an iterative algorithm to calculate the correct value (adjusted to a lower level if necessary) of the fictitious consumer has been implemented.

The dual situation is when the amount of water in the tank is less than the hourly demand. The simulator will point out that the water in the tank is not enough to supply the total hourly demand so it will be up to the operator's policy whether providing less consumers for

the whole hour or all the consumers for less than one hour.
The algorithm to calculate the Q_{in} for each tank also checks if there is water in the pipe adducting water to the tank and the tank is not filled when the answer is no. We considered this a major issue as, probably because tools like PSS-Sincal and most of the available simulators are supposed to be used for planning and not for emergency management, the commercial simulator raised the level of the tank anyway. This is why we detected this condition, set to 0 the fictitious consumer whenever the case and let the water level to decrease.
Non return valves (see Figure 7) have been managed too to avoid that the water tank supplies itself.
Last but not least, another overcome challenge is running the simulator when one (or more) tanks are empty, which in fact is the worst nightmare of all the simulators. In fact, when water level in tanks is 0, PSS-Sincal calculate the pressure in pipes as if at the node there was a column of water (generating pressure) whose height equals the node height, so consumers would continue to be supplied (which is not what happen in the real world). Having the capability to interface the simulator, the condition "no water in tanks" has been managed by forcing a temporary "out of service" condition so that pipes are properly emptied.

4. USE CASES

A number of incidents can destabilize the equilibrium of the water infrastructure. Among them, one of the most common is the assets failure (pumps, boosters, valves, electrical failure, etc) and to support the water distribution operator in setting their emergency plans it's very important to know the remaining storage capacity and, depending on the time of the day, how many hours are available for implementing the mitigation actions before the water tanks are empty and all the consumers run out of water. Water level in tanks has been identified as one of the main security metrics for water distribution networks.
In the following we consider: i) a fault or cyber-attack against the electrical grid which affects a critical pump and ii) combined faults and/or a cyber-attack compromising the treatment stations.

4.1. Pump Outage

As a first use case, we focused on a part of the network which, for geographical reasons, could be isolated and not supplied anymore if a pump failed (see Figure 8).
The main tank, marked as WT1 in Figure 8, has a volume of 2000 m^3 of water and supplies a few consumers as well as 5 tanks, each of them having their own consumers and supplying other tanks. The main tank is fed by the pump (marked as P in Figure 8) which adducts water from 2 reservoirs when the volume level is about half of the maximum one. System dynamic in normal operation is shown in Figure 9.
The pump, being at the same time an electrical and hydraulic device, is one of the elements of interdependency. In fact, problems in the electrical grid, for example some cyber anomalies on the ICT of its control system, can lead to the pumps being de-energised for some time, giving problems in supplying water to consumers. On the other side, unforeseen electrical loads (pumps, valves) due to the need to manage the water network under some cyber anomalies on the ICT of its control system, can impair the electrical grid.

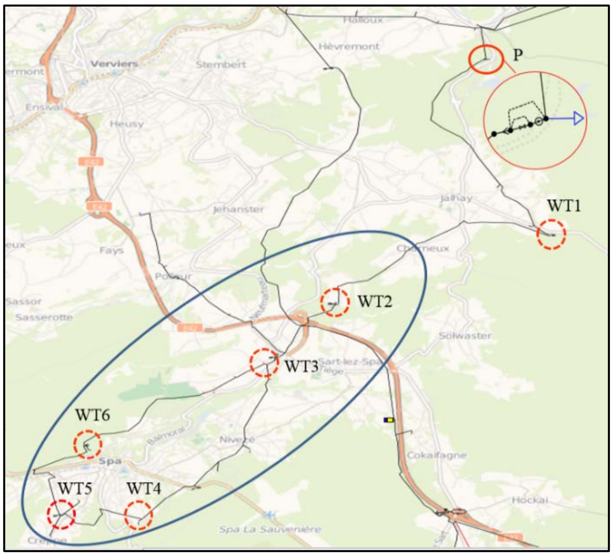

Figure 8: Sub-network which is affected by the Pump Outage

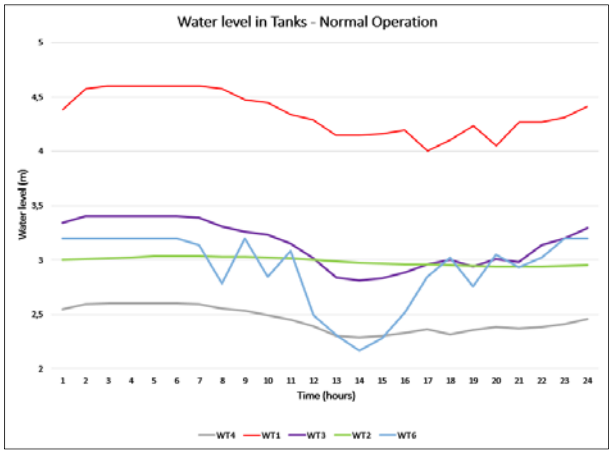

Figure 9: Dynamics of the sub-network in Normal Operation

The "out of service pump" condition can be generated both by a major physical fault in the electricity network and by a cyber-attack. In fact, pumps are usually installed in pairs in order to provide redundancy.
A likely scenario to justify this use case is an attacker who targets the pump P of the distribution water network that supplies the water storage facility WT1 to manipulate the water flow. This can be done:

- Indirectly, acting on the electrical grid, by cutting remotely power, the main pump of such a water storage systems stops for a long time in order to drain the water stored in WT1.

- Directly, acting on the water network control system to force the pump to stop in to order to drain the water stored as in the previous case or, even worse, to constantly pull water from the previous storage tank till to overcome the water storage level, also faking the related alarms.

The expected consequence is the tanks getting empty and the water distribution service being interrupted or needing to be handled with the emergency plan. Knowing the residual storage capacity and the forecasted demand is critical to optimize the emergency plans. The simulation output is discussed in Section 5.

4.2. Treatment station incident

Treatment station is a critical element of the water distribution system. Intentional human actions to make a shift in the water quality is by far one of the worst threats but even a severe electrical failure can lead to a simple treatment station incident and affect a huge number of consumers.

In fact, after being collected (in our case from a dam), raw water is led to treatment stations to be converted into drinkable water in line with the quality requirements. This process is executed over multiple stages (filtration, remineralisation, disinfection, pH adjustment) which involve the operation of electrical devices such as pumps and valves, as well as sensors for chemical adjustment and check. If water is not in line with quality requirements, water cannot be input in adduction pipelines.

As already stated in Section 2, in the water distribution network there are 2 reservoirs with treatment stations. The simulator allows us to study what happens when one or both the reservoirs are disconnected including details at each node about the reservoir supplying the water (where the water comes from).

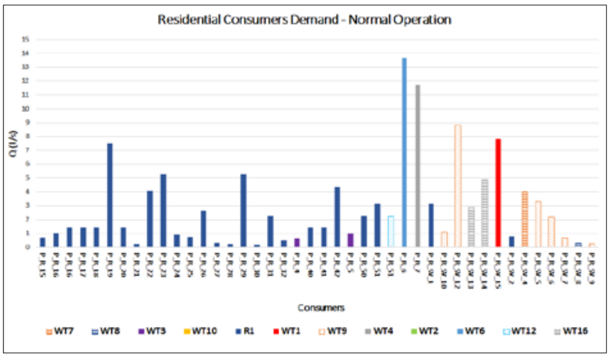

Figure 10: Residential Consumer Demand and Source of Water Supply

Figure 10, Figure 11 and Figure 12 show the demand of all the consumers, while the colour and pattern of the bars identifies the tank or the reservoir which supplies the water in normal conditions.

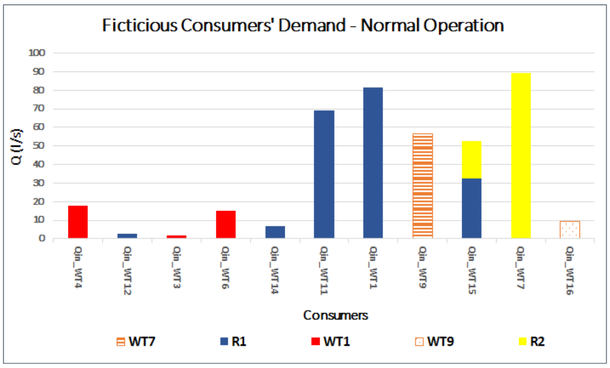

Figure 11: Fictitious Consumer Demand and Source of Water Supply

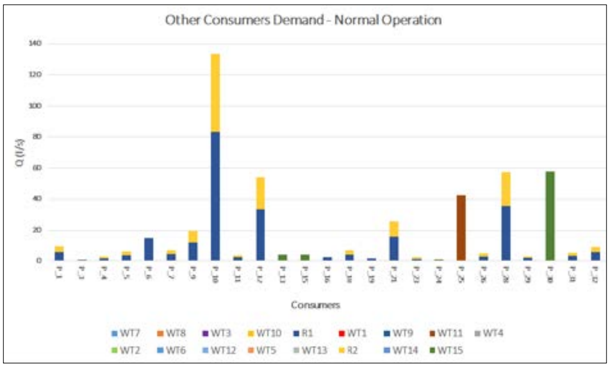

Figure 12: Other Consumer Demand and Source of Water Supply

It can be noticed that, in normal conditions most of the consumers are supplied exclusively by one reservoir (or one tank supplied by the reservoir) or the other, while some of them (showing 2-colours bar) get the water from both. The length of the coloured sections (Figure 11 and Figure 12) shows the percentage of water coming from one reservoir or the other.

5. SIMULATION RESULTS

Simulation of the use cases described in the previous Section 4 has been carried out for 24 hours with the purpose of computing the remaining storage capacity and the time the tanks will take to get empty. It's worth pointing out that, although as detailed above the out-of-service condition can be generated by different types of malfunctioning, the simulation has been carried out putting out-of-service the elements affected by the scenario (the pump P in the pump outage scenario and the reservoirs in the treatment station incident), independently by the specific occurred cause. The simulations start at time 00:00 and end at time 23:59.

A simulation in normal conditions has been carried out in order to make a comparative check that the behaviour of the water network and the dynamic of the tanks are as expected.

Simulation output shows that the service is delivered to water consumers for the whole timeframe, pressure and velocity of the water within the pipes result within the normal range and water level in all the tanks is always above the threshold (see Figure 13).

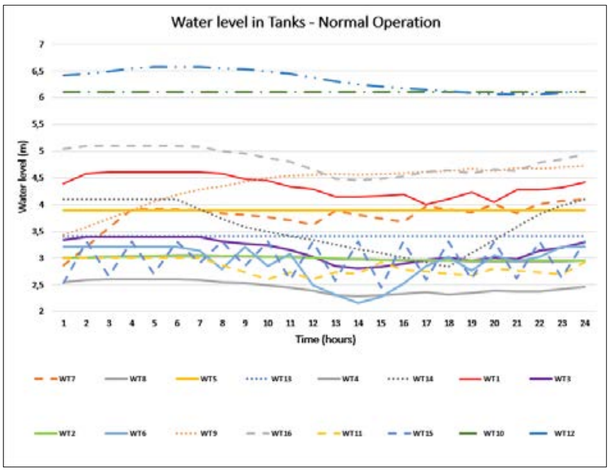

Figure 13: Water Level in all the tanks of the network in Normal Operation

Figure 13 shows different “shapes” of the water level in the different tanks, due to the settings of the filling pumps. The quicker water is pumped into the tank with respect to the volume, the steeper is the profile, like it happens for the water tank WT15 (represented by a blue dotted line in Figure 13).
For some tanks water level is almost constant due to their very small load.
Figure 15 shows what happens in the tanks as soon as the pump P is out of service (time 00:00).
The water level in the WT1 tank (red line) start decreasing immediately while supplying its own consumers and keeping in service all the tanks downstream which operate normally until the WT1 tank become empty. As soon as the WT1 is empty all the tanks downstream start emptying, each one according to the demand of their own consumers. In fact, due to the specific elevations, the different tanks do not supply each other (see Figure 14).

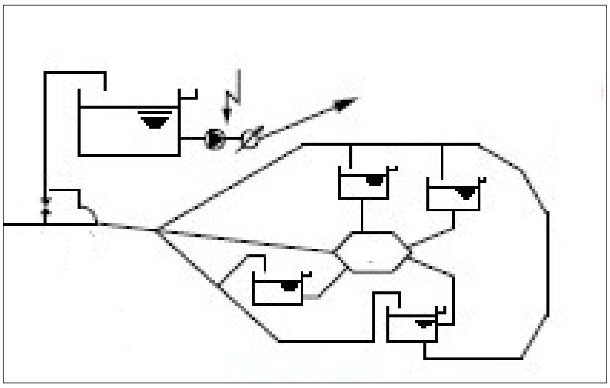

Figure 14: Water Tanks loop supplied by WT1

As shown in Figure 15, the most critical tank after WT1 is WT6 (light blue line) as it has a quite high demand compared to the tank capacity, whose autonomy is just 3 hours since the WT1 tank is out of water. Six hours later, WT3 will be empty too and within 14 hours all the consumers except those supplied by WT2 will run out of water.

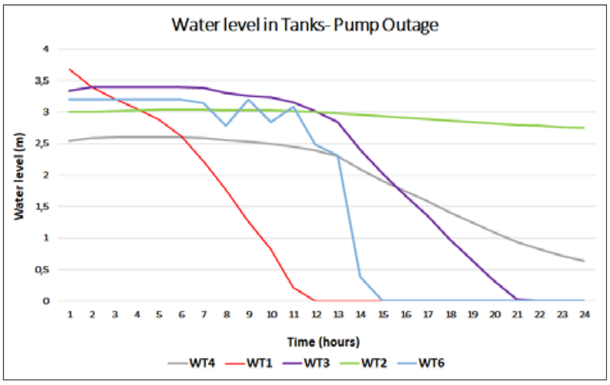

Figure 15: Water Level in the tanks affected by the Pump Outage

In this simulation, the WT1’s 12 hours of autonomy shown in Figure 15 is based on the assumption that the initial value of the water level was 3.70 m. This depends on the filling policy and on the time the fault happens. On the other hand, as the pump under analysis is quite a big load for the electrical network, it cannot be switched on at any time, so we recommend to keep the tank as much filled as possible (for example in the night when the electrical load is lower).
Last, it’s worth pointing out that the total demand by WT1 is 150 m^3/h and – having water trucks with a capacity of 30 m^3 – the emergency plan should ensure that it is possible to fill and deliver up to 5 trucks per hour to maintain the service to all the consumers, and checking the control system to notify as soon as the pump is faulty even if it is not running so that it is possible to have a pre-alarm phase (while WT1 is not empty yet) which is 12 hours long and can be used to put in place a mitigation strategy, otherwise the alarm on the water level will just raise the emergency plan.
In order to handle the scenario about the treatment station incident, three simulations have been carried out. As reported in Section 2, the water network under study is supplied by 2 reservoirs which can be considered infinite sources as long as the associated treatment station is properly working. Assuming that only one treatment station at the time is out of service (due to a problem on the electricity supply putting the treatment station to an halt or due to an alarm on water quality raised by the SCADA system) then preventing the reservoir from supplying the distribution network, two simulations have been carried out in order to evaluate the impact on the network. Differently from what expected, as far as it concerns the service delivery and the water level profiles in the tanks, there are no relevant changes. However, PSS-Sincal has a node-tracking functionality that allows to know which tank/reservoir supplied the water at each node and with which percentage. Therefore it has been possible to notice that although the consumers don’t notice any change, the whole amount of water is supplied by the

working reservoir. A double check with the operator confirmed that the treatment process is quick enough for the working reservoir to supply all the consumers, although the reverse flow in the pipe could give some quality water problems that need to be investigated further and managed depending on the emergency situation. Figure 16 and Figure 17 show that in one simulation the water was supplied by reservoir 2 (in yellow) and in the second by reservoir 1 (in blue).
Simulation results show that the water infrastructure is robust with respect to faults. In fact, if just one of the 2 reservoirs is shut down (for example due to a problem to the treatment station) simulation output shows that the whole system is taken up by the other reservoir.

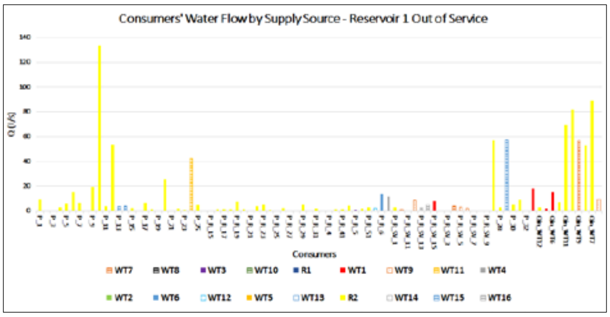

Figure 16: Reservoir 1 Out of Service

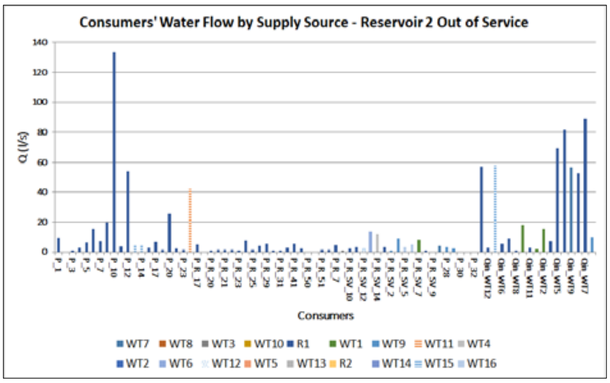

Figure 17: Reservoir 2 Out of Service

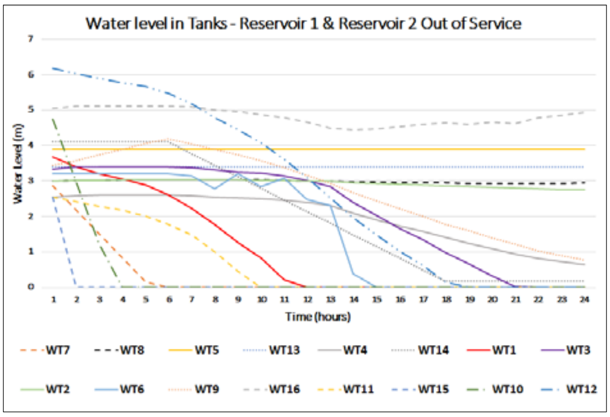

Figure 18: Water Level in tanks when both Reservoirs are Out of Service

However, a massive widely spread set of faults (likely to happen in case of a severe natural hazard or in a coordinated cyber-attack) will jeopardize the water system. Figure 18 shows the water level in tanks when both reservoirs are out of service. As it can be seen, some consumers are out of service after only a couple of hours and within 18 hours most of the consumers are not supplied any more. In fact, those tanks who are not heavily affected by the outage are connected to consumers with very low demand. Unfortunately, geographical constraints do not always allow to reconfigure the network to use the water in those tanks.

6. CONCLUSIONS

Critical Infrastructures provide essential services that support economic prosperity, governance and quality of life. Although as citizens we all are used to their continuous and reliable functioning, socio-economic as well as political changes in the last 20 years are about to compromise the quality of the service provided. Therefore, besides identifying the vulnerabilities of the single components, a system-wide assessment of the vulnerabilities of a network is strongly recommended. To this purpose, within the scope of the EU Horizon 2020 project ATENA and with the Belgian water operator's support, a water distribution simulator has been developed. Taking into account the physical topology and the behaviour of the system, it is useful to evaluate the impact of multiple faults (due to the most diverse reasons, from interdependency to natural hazards to malicious or even cyber-attacks) and therefore to set up, validate or improve the operator's emergency plans.
Although the most accurate available simulators are the commercial ones, tests showed that – due to the different scope – some functionalities needed to be refined. Being able to interface the commercial simulator was a critical factor for success and our simulator – enhancing PSS-Sincal by Siemens with ad hoc modules – inherits its benefits in terms of reliability and operator's trust in the results but, at the same time, overcomes its restrictions.
The improved capabilities have been tested to support emergency plans in a couple of use cases which have been detailed, one about the outage of a critical pump, the other about a treatment station incident. In both cases the main metrics for the operator are the water level in tanks and the number of hours they can still provide the service. The developed simulator was able not only to calculate them for the different tanks, but also – being geo-localized – made it possible to identify the timing of all the impacted areas.
Simulation results confirm the common sense of robustness for the water system, but also the worries of the different stakeholders about the impact a combined attack would have, affecting a huge number of consumers missing a vital service.
The authors think this work is worthy because the size of the modelled network, the quality of the data and the improved capabilities of the developed simulator, which will be soon connected and synchronized to the electrical simulator under development.

Simulation output in normal operation looks realistic therefore future work include both i) a results validation phase aimed at comparing the simulation results with the measurement coming from sensors deployed on the infrastructure (acquired from the SCADA system) in order to evaluate their gap and ii) a simulation campaign aimed at identifying interesting and counterintuitive use cases which it is worth setting up an emergency plan about.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Program, under Grant Agreement no 700581 (ATENA project) www.atena-h2020.eu

REFERENCES

ATENA Horizon 2020 project, 2016-2019. Advanced Tools to assEss and mitigate the criticality of ICT compoNents and their dependencies over Critical InfrAstructures. Available from http://www.atena-h2020.eu

Balducelli C., Di Pietro A., Lavalle L., Vicoli G., 2007. A Middleware Improved Technology (MIT) to Mitigate Interdependencies between Critical Infrastructures. In: de Lemos R., Di Giandomenico F., Gacek C., Vieira M. ed. Architecting Dependable Systems V, 28-51.

Boano F., Scibetta M., Ridolfi L., Giustolisi O., 2015. Water distribution system modeling and optimization: a case study. Procedia Engineering 119, 719-724.

Dakin R., Newman R., Groves D., 2009. The case for cyber security in the water sector. J. Am. Water Works Assn. 101.

EU FP7 CockpitCI project. Available from: https://cordis.europa.eu/result/rcn/173345_it.html

Horta R., 2007. The city of Boca Raton: a case study in water utility cyber security. J. Am. Water Works Assn. 99.

Leclaire R.J., Pasqualini D., Dreicer J.S., Toole G.L., Urban N.M., Bent R.W., Mcpherson T.N., Hengartner N.W., 2014. Infrastructure Modeling: Status and Applications. In: Linkov I., ed. Sustainable Cities and Military Installations, 391-427.

McPherson T.N., Burian S.J., 2005. The Water Infrastructure Simulation Environment (WISE). Project. Presented for the EWRI'05 World Water & Environmental Resources Congress.

Pederson P., Dudenhoeffer D., Hartley S., Permann M., 2006. Critical Infrastructure Interdependency Modelling: A Survey of US and International Research. Idaho National Laboratory, Idaho Falls, Idaho.

Perelman L. and Amin S., 2014. A network interdiction model for analyzing the vulnerability of water distribution systems. Proceedings of 3rd Int. Conf. on High Confidence Networked Systems, pp. 135-144. New York.

PSS-Sincal. Available from: www.siemens.com

Ramesh Dr. H., Santhosh L., Jagadeesh C.J., 2012. Simulation of hydraulic parameters in water distribution network using EPANET and GIS. Proceedings of Int. Conf. on Ecological, Environmental and Biological Sciences (ICEEBS'2012), Dubai.

Rossman L.A., 2000. EPANET 2 Users Manual. Available from: www.epa.gov/water-research/epanet#downloads

Taormina R., Galelli S., Tippenhauer, N.O., Salomons E., Ostfeld A., 2017. Characterizing cyber-physical attacks on water distribution systems. J. Water Resour. Plann. Manage. 143.

AUTHORS BIOGRAPHY

Tatiana Patriarca received the degree in Communication Engineering in 2008 and the Ph.D. in Information and Communication Engineering in 2012 from the University of Rome. She is currently researcher at ENEA (Italian national Agency for New Technologies Energy and Sustainable Economic Development) and her research activity is in cyber security field of Critical Infrastructures. She works in ATENA Horizon2020 EU project and she has worked in CockpitCI EU FP7 project. Previously, she worked in Telecom Italia, main Italian ICT group and provider of telecommunication services, and in Ericsson, provider of services, software and infrastructure in ICT technology for telecom operators. She is co-author of several scientific papers.

Luisa Lavalle received her master degree in Electronic Engineering from University of Palermo in 2000. After 3 years in the R&D of a multinational company in the Telco domain, she has been working at ENEA (Italian national Agency for New Technologies Energy and Sustainable Economic Development) since 2003 and her research activity is on Critical Infrastructure Protection, focusing on decision support systems algorithms and impact assessment. She has been involved in several European Project (SAFEGUARD, IRRIIS, SEESGEN-ICT, CIPRNET, ATENA) and is co-author of several scientific papers.

Ester Ciancamerla received her degree in Nuclear Engineering from University of Rome in 1978. Since her degree, she has been working at ENEA, as scholarship holder and researcher. In remote past, major experience has been gained dealing with system validation, software verification plans, software test methodologies, tools and environments for computer based systems in nuclear, avionics and railway fields. Her current research interest is on modelling methods and tools for dependability/survivability evaluation of networked systems. She has acted in the frame of several research programs, funded by Italian research organisations and by European Union; among the most recent SINERGREEN, an Italian project and CockpitCI, MICIE, SHIP, ISAEUNET, SAFETUNNEL, IRRIIS EU Projects. He has authored and co-authored more than 90 papers for International Journals and Conferences Proceedings.

Olivier Hautier achieved his Engineering Master in 1996. He counts a 15+ years of experience in the field of Process Automation and Industrial. In 2014 Olivier became Senior Technical Consultant and Project Manager as a freelance for the Société Wallonne des Eaux wherein he leads the development and the deployment of a the centralized SCADA application that covers the 1000+ production sites in Wallonia - Belgium.

Bernard Daulne received the degree in Industrial Engineering in 1992 from the Zenobe Gramme Institute of Liège. He is currently expert in industrial computing at SWDE (Société Wallonne des Eaux) where he implements a supervisory system which include the standardization of each CIM layer. He has a many years' experience in electromechanical infrastructures studies and in water production networks management.

Machine Learning for Approximated Sensors for Marksmanship Training

Keith Brawner[(a)]
[(a)] United States Army Research Laboratory, Human Research and Engineering Directorate

[(a)]Keith.W.Brawner.Civ@mail.mil

ABSTRACT
All people in the military must be proficient on the basics – to shoot, move, and communicate. Basic Rifle Marksmanship (BRM) is required for both non-commissioned and enlisted Warfighters in all branches of military service, with training on BRM skills being conducted in a series of "dry fire", simulation, and live drills. In all phases of training, Warfighters receive instruction on the four fundamentals of shooting: breathing, body position, sight picture, and trigger squeeze. Within simulation, this training is conducted in a 1:4 to 1:8 range; one instructor per 4-8 students. While realtime individualized feedback is a goal of instruction, it is not a reality, as instructors must attend to the needs of many students. In an effort to aid instructors in providing valuable individualized feedback, a tutoring system was developed which automatically diagnosed novice performance when compared to experts. This system was used to diagnose novice performance with extensive physical sensors applied to the weapons. This work investigates whether machine learning can aid in the diagnosis of the novice performance, without the physical sensors, and analyzes the degree to which the sensors are necessary.

Keywords: Marksmanship, Intelligent Tutoring Systems, Machine Learning

1. INTRODUCTION TO MARKSMANSHIP INSTRUCTION

Marksmanship has been described as a complex psychomotor skill demanding high physical and mental coordination. The end goal of the instruction is the ability to hit a target with a bullet launched from a rifle at a reasonably far distance (300+ meters). The psychomotor aspects of marksmanship demand a control of the breathing, body, trigger, and the sight picture of the rifle. In the US Army, Basic Rifle Marksmanship (BRM) training consists of instruction on these fundamentals of marksmanship, in addition to grouping and zeroing an M16-/M4-series rifle, transitional firing (United States Army, 2003), and a culminating qualification event. Traditionally, this training includes education in the classroom, practice within a simulated environment (ie. the Engagement Skills Trainer (EST)) and a live fire exercise on physically outdoor ranges. Of particular interest to this work is the EST simulation portion of the standard training. As a simulation, the EST captures significant amounts of information about the learner and their actions. The capture of this data in simulation performance allows for the potential injection of Intelligent Tutoring System (ITS) technology which can automatically instruct the learner. If the system has the data on the learner's actions, it is possible to act upon it and instruct the learner.

Based on the data gathered from learner patterns, learner behaviors, and populations of leaners, ITSs have the general goal of automating and augmenting human instruction. Fundamentally, one-to-one human-to-human instruction has been shown to be the most effective form of instruction (Bloom, 1984), but for practical purposes there are limited instructor resources. In practice, marksmanship training has between four and eight students per instructor at either the live or simulated ranges. Similar to a classroom, students practice their skills continuously, rendering it practically impossible for instructors to give individualized feedback at on shots or groups (5 shots). While individual human instruction is effective, it is simply not practical for continuously occurring tasks. ITS technology, however, makes individual instruction a possibility, as a computer can monitor many individuals simultaneously. Further, in a study where human checklist-based diagnosis of novice marksman performance was used, the authors noted a general increase in performance over baseline methods (Chung, Nagashima, Espinosa, Berka, & Baker, 2009), which gives credibility to the current work in computationally automating such assessments. The current work additionally extends the diagnosis to additionally include the provision of feedback.

2. PRIOR WORK

2.1. Initial data collection and models

In order to give individualized feedback on shots and groups, an initial study was designed for the purpose of making models of expert performances. This study involved the collection of many shots and shot grouping from eight experts. Despite that there were few experts collected, there was evidence that the experts performed similarly to each other; there was model alignment between the extracted features. These similarities were further aligned to the BRM manuals and instructions (Goldberg, Brawner, Amburn, & Westphal, 2014). In short, we observed that experts fired their weapons in manners similar to each other, and in a manner similar to what is prescribed in the BRM manual. This gives some weight to the idea that such models could be used to diagnose novice performance, as the BRM manual is the baseline against which novices are compared by human expert instructors.

We used this model of expert performance to create initial error-based domain models for three of the four primary instructions of BRM: Trigger Squeeze, Breathing, and Weapon Orientation. The fourth BRM fundamental - body position – was not created, as there was not a weapon-based sensor which could be easily applied to create such a model. The created models which were created were based on the data sources for a trigger pressure sensor, breathing strap, and high-resolution orientation sensing mechanism. A picture of the sensor-ized weapon and breathing strap is shown in Fig. 1. These models were cross-validated at the individual expert level (Goldberg et al., 2014). After observing general model performance on the SCATT marksmanship system, shown in Fig. 1, similar models were created for the FATS M100 system (Goldberg, Amburn, Ragusa, & Chen, 2018). The M100 models were the models used throughout the rest of this work.

The basic procedure for the collection of data for expert models was that each of eight experts in sequence arrived to a system which was pre-zero'd, fired as many shot groupings (5 shots) as they could within a timeframe, and had self-administered breaks to account for physical and eye fatigue. The experts fired shots in four conditions – either standing or kneeling and either with or without gear. The Generalized Intelligent Framework for Tutoring (GIFT) was used for both the initial data collection and initial user study (Sottilare, Brawner, Sinatra, & Johnston, 2017).

2.2. Initial User Study

Although a study of the behavior of experts is intellectually interesting, the purpose of making models of expert behavior is to *use* them to diagnose novice performance. The purpose of diagnosing novice performance is to identify the weaknesses and apply specific training to the subject areas; giving individual feedback. The end purpose of this feedback is the outcomes of training, such as performance improvement and behavior change. An unpublished study was run to determine the effectiveness of the total intelligent tutoring system, BRM system, and their combination with respect to their effect on novice performance. Naturally, novice marksman made many mistakes. These mistakes were noted by the system and afterwards used to trigger feedback. The system diagnosis of novice mistakes was logged significantly for later post-processing, in current work. The educational/performance effect of the total system is currently in publication preparation by other authors.

The end result of the system is to make decisions on which individuals errors. Those decisions are truncated from the models – they are made on data post-processed after it is run through the models. From a "black box" systems perspective, a system which produces identical outputs from inputs the same system, which brings us to the current work – recreating a similar system with dramatically fewer internal working parts. This begs the question "do we need the full system of sensors and models, or can we model the behavior of the resultant system instead?" Can the internal working which rely on sensors and experts be replaced with algorithmic solutions which closely model these intermediate processes?

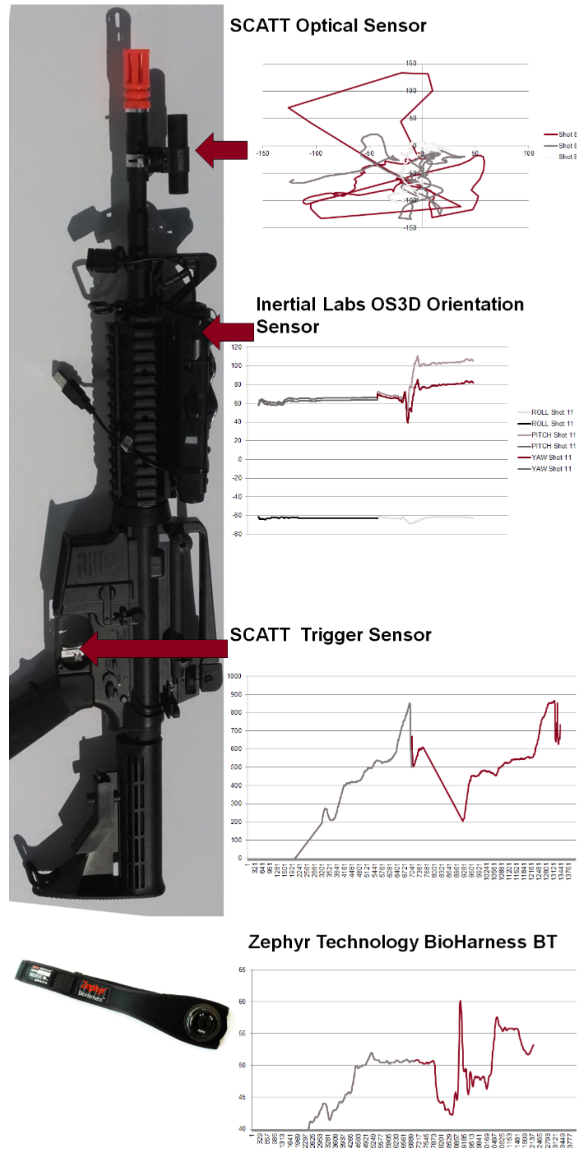

Fig. 1: Initial Experimental Configuration (Amburn, Goldberg, & Brawner, 2014) using the SCATT system. Further models using the proprietary FATS M100 system created in near-identical fashion in further work.

3. MACHINE LEARNING MODELING

The above experiments present a picture of initial system setup, data collection from experts, using this data to make models, and using those models to diagnose novice performance. This work is summarized in Fig. 2. The next steps for this work involve automating the diagnosis with a minimal number of sensors, as shown in Fig. 3.

It is clear that the patterns of overall barrel movement and other marksmanship fundamentals are correlated. As an example, consider Fig. 4, which shows two sensors data streams 5 seconds prior to the individual shot. A slow trigger squeeze is associated with very little barrel movement, while sharp trigger jerks have an overall effect on the barrel movement. It is clear that the patterns are correlated, the process of creating machine learning models answers the question of "to what degree?" and the follow-up questions of "if they are highly correlated, can we eliminate sources of data?" and "what degree of error will this generate?".

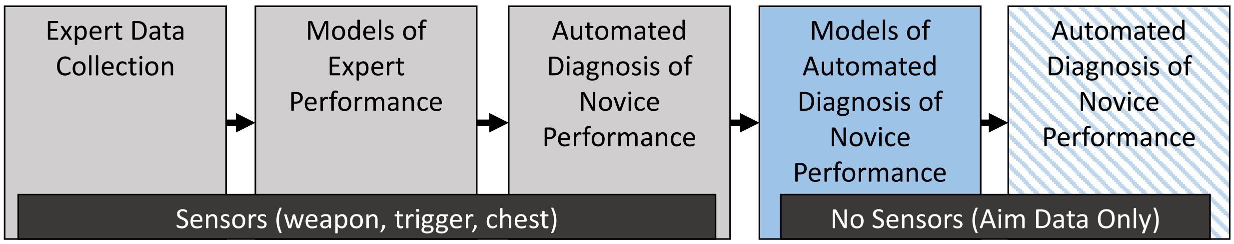

Fig. 2: Model of the total process of data collection and analysis

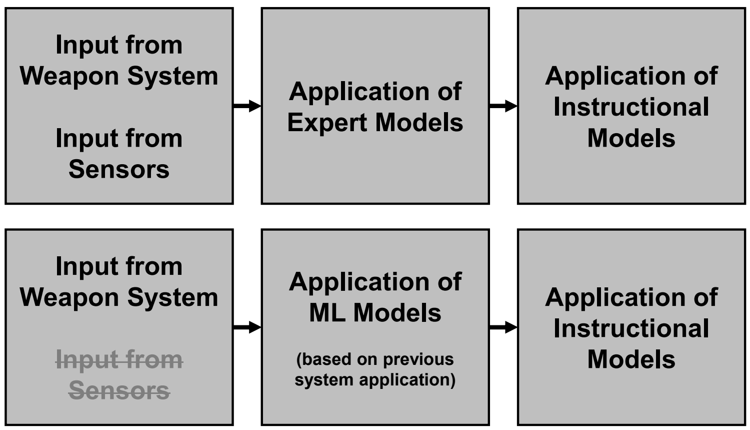

Fig. 3: Model of the total process of machine learning model replacement

Proceedings of the International Defence and Homeland Security Simulation Workshop 2018
ISBN 978-88-85741-12-6; Bruzzone and Sottilare Eds.

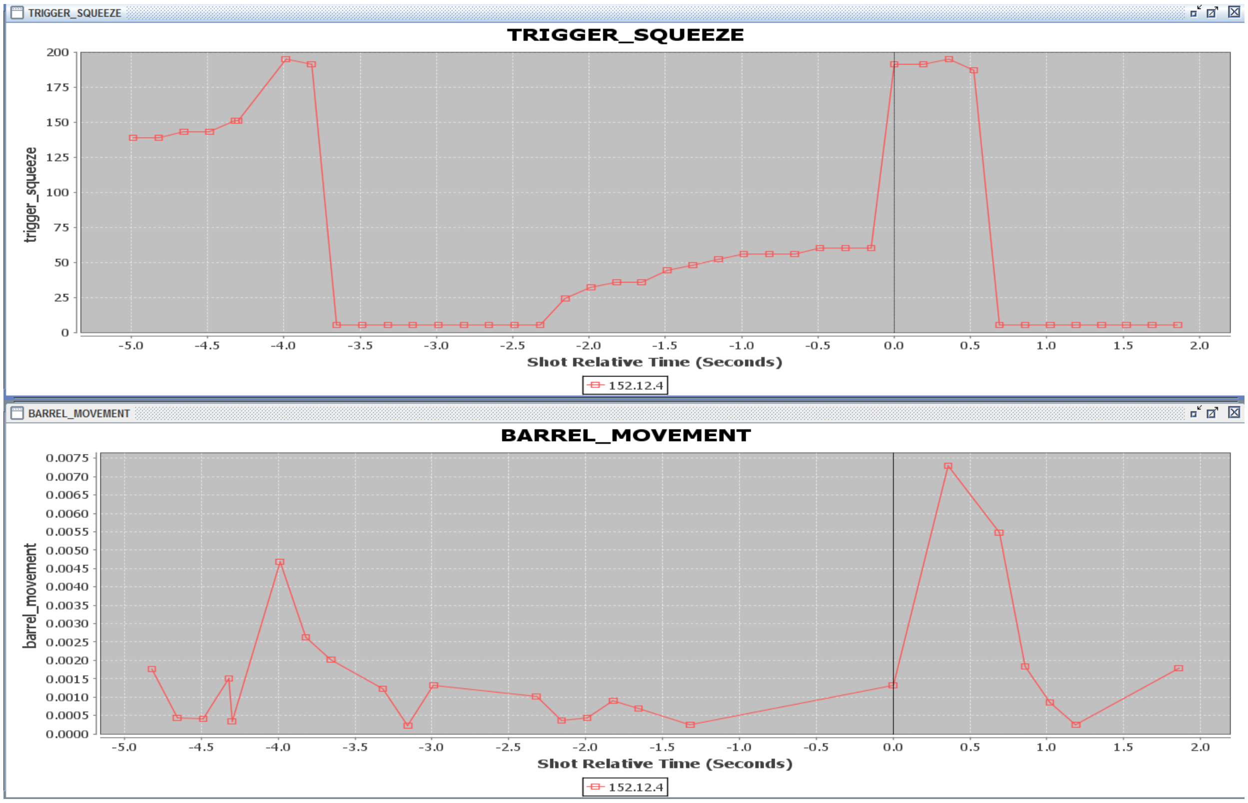

Fig. 4: Example of data output. The reader should note that slow trigger squeeze indicates small barrel movement.

A list of the features which were used for the machine learning models is summarized within Table 1. Generally, these features correspond to the screen positions, movements, summaries of the movement, and statistical features and Fast Fourier Transforms (FFT) for various time windows (1, 3, 5 seconds). These types of features were used because this is what saw success in the creation of expert models, which were based roughly on military instruction and procedure.

In order to create the machine learning models, RapidMiner was used (Hofmann & Klinkenberg, 2013). The settings were configured to use the system diagnoses as the labels for machine learning models of various types, as shown in Fig. 3. As an example, a student took many actions in regards to the marksmanship fundamentals, took a shot, and the system diagnosed the error as being related to trigger pull. Whether or not this is the correct diagnosis for the novice, this is provided as the label to RapidMiner models as the ground truth – the machine learning system is attempting to *replicate* the behavior of the real system, not to *improve* upon it. The improvements in using such a system are related to the cost improvements from the lack of sensors (bodily or otherwise). 10-Fold cross-validation at the shot level (10% of total data withheld, randomly selected, for testing) was used with a variety of models.

The machine learning methods used to build models include a Naïve Bayesian implementation (NaiveBayes), a one-level decision tree (DecisionStump), a multi-level decision tree (DecisionTree), a linear regression decision tree model (GradientBoostedTree), a randomly seeded forest of decision trees (RandomForests), a baseline neural network methods (Artificial Neural Networks; ANN), and a Support Vector Machine (SVM) implementation. Metrics are captured on each of these metrics for accuracy, precision, and recall.

ScreenX
ScreenY
BarrelMovFFTMag
Barrel Mid point
BarrelMovSTD
BarrelMovMean
Screen X 1s
Screen Y 1s
Screen X 3s
Screen Y 3s
Screen X 5s
Screen Y 5s
Screen X 10s
Screen Y 10s
time1 std
time1 mean
time1 mid point
time3 std
time3 mean
time3 mid pont
time5 std
time5 mean
time5 mid pont
time10 std
time10 mean
time10 mid pont
Barrel Movement X Barr1secondX
Barrel Movement Y Barr1secondY
Barrel Movement(mag) Barr1secondMag
Barrel Movement X Barr3secondX
Barrel Movement Y Barr3secondY
Barrel Movement(mag) Barr3secondMag
Barrel Movement X Barr5secondX
Barrel Movement Y Barr5secondY
Barrel Movement(mag) Barr5secondMag
Barrel Movement X Barr10secondX
Barrel Movement Y Barr10secondY
Barrel Movement(mag) Barr10secondMag
Barrel Movement X
Barrel Movement Y
Barrel Movement(mag)
Barrel Movement(magFFT) Barr1secondFFT
Barrel Movement(magFFT) Barr3secondFFT
Barrel Movement(magFFT) Barr5secondMagFFT
Barrel Movement(magFFT) Barr10secondMagFFT

Table 1: Features for the machine learning model prediction.

4. RESULTS

Presented in Table 2 are the results of applying the previously mentioned machine learning models. Generally, this is thought of as an "out of the box" implementation which represents a starting point from which to build. No particular effort for parameter tuning of advanced model-making was undertaken; this is left as an exercise for technology transition. This research seeks to answer the question of whether or not this is a worthwhile task.

The models are assessed using standard measures of model fit: accuracy, as a percentage correct diagnosis, precision, as a measure of identifying the relevant datapoints, and recall, as a measure of insurance against missing important datapoints. These measures of model fit should be suitable for an individual task which is not particularly unbalanced. This dataset is not badly unbalanced in the dimension of error category – novice marksmanship practitioners made mistakes on all areas of marksmanship.

Barrel	**Accuracy**	**Precision**	**Recall**
DecisionStump	61.84%	61.84%	100.00%
DecisionTree	61.84%	61.84%	100.00%
GradientBoostedTree	*98.89%*	*99.09%*	*99.11%*
RandomForests	61.84%	61.84%	100.00%
ANN	76.49%	79.23%	84.00%
SVM	76.49%	79.23%	84.00%
Trigger	**Accuracy**	**Precision**	**Recall**
DecisionStump	66.48%	66.29%	100.00%
DecisionTree	67.14%	66.73%	100.00%
GradientBoostedTree	*98.89%*	*99.06%*	*99.25%*
RandomForests	65.92%	65.92%	100.00%
ANN	77.10%	79.08%	88.74%
SVM	65.76%	66.00%	99.15%
Breathing	**Accuracy**	**Precision**	**Recall**
DecisionStump	84.10%	84.09%	100.00%
DecisionTree	84.52%	84.45%	99.99%
GradientBoostedTree	*90.14%*	*90.16%*	*99.09%*
RandomForests	84.06%	84.06%	100.00%
ANN	90.18%	90.59%	98.56%
SVM	84.06%	84.06%	100.00%

Table 2: Results of the machine learning models as applied to the individual problems.

The above results of various machine learning measures over the dataset contain indications about the type of classification problem presented to the algorithm. As an example, ANNs typically have problems in dealing with classification spaces which are not linearly separable. Given the above results, there is reason to believe that the underlying data streams from each of these sources is approximately linear discontinuous in classification space, as observed in the high performance of GBT/ANN/SVM implementations, with special emphasis on GBT implementations, which explicitly model linear discontinuous classification spaces.

Having a linear discontinuous classification space is logical, as the sensor feeds stimulating the real system and are eventually classified as performance levels of either "above expectation", "at expectation", "below expectation", or "unknown". Logically, these labels are discontinuous by their nature – a small variation in target accuracy or trigger pull shifts someone from one of these categories to the other one. Given this type of input (e.g. sensor streams), output (e.g. finite bins), and reasonably direct relationship between items, it is relatively unsurprising that a bounded form of linear regression models is the best performing. Gradient Boosted Tress (GBT) are exactly this type of discontinuous linear regression model, and thus have exceptional performance.

The accuracy in the above table is startlingly accurate (~100%) when modeled according with a model which is close to the model of the underlying phenomenon, and moderately accurate when incorrectly modeled (~80%). This gives some indication that the system can be safely modeled without the need for the sensor data streams – much of the data underpinning the sensor data stream is buried within the aim trace data, in manners unknown to the experimenters. Naturally, the patterns of behavior within the aim trace data would not have been able to be discovered without the sensor data to begin with.

Further, due to the programming of the system, the inaccuracies are relatively more forgiving when considering *actual* tutoring actions. The tutoring actions of the actual system have a specified order to tutoring (breathing, then trigger, then body position, then sight picture), as shown in Fig. 5. Only one marksmanship fundamental is tutored at a time in order to limit the amount of information presented to the learner. As such, an inaccuracy in a downstream system is irrelevant; it doesn't matter what the accuracy of the barrel predictor if the end result is going to be trigger feedback. Additionally, short-duration feedback presented to the trainee after they miss the target isn't a large detractor; there is little penalty for giving short feedback. Specifically, a learner receiving instruction on how to control their breathing, when the classification was incorrect, is not highly detrimental to overall training. Similarly, in practice, the system is configured such that learners are only given feedback when poor performance is observed; hitting the target consistently results in no feedback being given. Given the above system considerations, even an 80% accuracy rate may be acceptable for production, considering that this is falsely reflective of actual system error and saves real dollars in the short term through the elimination of sensors and the time that it take to configure them. Given the results presented in Table , there is some indication that this type of model can be safely transitioned to practice.

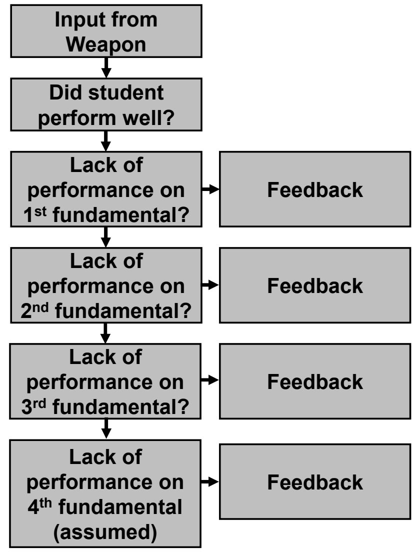

Fig. 5: Design of feedback delivery based on marksmanship mistakes. Note that an 80% error in the 3rd classification area will only be presented to the student in a fraction of the total instances; an 80% model error rate represents a *maximum* of an 80% system error rate; actual error is much lower.

5. CONCLUSIONS AND FUTURE WORK

In conclusion, there is some significant reason to believe that the sensors are not strictly necessary for diagnosing marksmanship performance deficiencies within an intelligent tutoring system. Answering the initially posed research question: "Can the internal working which rely on sensors and experts be replaced with algorithmic solutions which closely model these intermediate processes?", the answer is "yes". The accuracies within the presented results table are startlingly accurate, and even if they were not, it wouldn't be much of a consideration for the live production system. Given the conclusion that the total system can be modeled with reasonably high accuracy, which is higher when considering the "one feedback at a time" model, there is little reason to include the sensors in practice – clever modeling can eliminate the requirement of sensors. Naturally, this doesn't obviate the total collection work – the end system must still collect enough novice data in order to make the sensor-free models; the sensor-free models cannot be calculated in the absence of labeled data from system performance, nor are they easily simulated.

ACKNOWLEDGMENTS

The author would like to acknowledge Benjamin Goldberg, Tommy Goris, and Cadet Shane Ko for the initial data collection, initial data analysis and subsequent data analysis. This project was a significant collaboration over time.

REFERENCES

Amburn, C., Goldberg, B. S., & Brawner, K. W. (2014). *Steps Towards Adaptive Psychomotor Instruction.* Paper presented at the FLAIRS Conference.

Bloom, B. S. (1984). The 2-Sigma Problem: The search for methods of group instruction as effective as one-to-one tutoring,. *Educational Researcher, 13*(6), 4-16.

Chung, G. K., Nagashima, S. O., Espinosa, P. D., Berka, C., & Baker, E. L. (2009). An Exploratory Investigation of the Effect of Individualized Computer-Based Instruction on Rifle Marksmanship Performance and Skill. CRESST Report 754. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.

Goldberg, B., Amburn, C., Ragusa, C., & Chen, D.-W. (2018). Modeling expert behavior in support of an adaptive psychomotor training environment: A marksmanship use case. *International Journal of Artificial Intelligence in Education, 28*(2), 194-224.

Goldberg, B., Brawner, K., Amburn, C., & Westphal, M. (2014). *Developing Models of Expert Performance for Support in an Adaptive Marksmanship Trainer*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Hofmann, M., & Klinkenberg, R. (2013). *RapidMiner: Data mining use cases and business analytics applications*: CRC Press.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Retrieved from

United States Army. (2003). Rifle marksmanship M16A1, M16A2/3, M16A4, and M4 carbine (FM 3-22.9). *Washington, DC: Headquarters, Department of the Army*.

Keith Brawner, Ph.D. is a researcher within the Learning in Intelligent Tutoring Environments (LITE) Lab within the U. S. Army Research Laboratory's Human Research & Engineering Directorate (ARL-HRED) in Orlando, Florida. He has 12 years of experience within U.S. Army and Navy acquisition, development, and research agencies. He holds a Masters and PhD degree in Computer Engineering with a focus on Intelligent Systems and Machine Learning from the University of Central Florida. His current research is in machine learning, active and semi-supervised learning, ITS architectures and cognitive architectures. He manages research in adaptive training, semi/fully automated user tools for adaptive training content, and architectural programs towards next-generation training.

HOLOVULCANO: AUGMENTED REALITY SIMULATION OF VOLCANIC ERUPTIONS

Ali Asgary

Associate Professor, Disaster & Emergency Management, School of Administrative Studies, York University, Toronto, Ontario, Canada

asgary@yorku.ca

ABSTRACT
This paper describes an interactive holographic simulation of volcanic eruption. The aim of the project is to use Augmented Reality (AR) technology to visualize different volcanic eruptions for public education, emergency training, and preparedness planning purposes. To achieve this goal, a 3D model of the entire Vulcano Island in Italy has been created using real elevation data. Unity game engine and Microsoft Visual Studio have been used to develop HoloVulcano augmented/virtual reality simulation application. The current version of HoloVulcano simulates normal and unrest situations, single and long lasting Vulcanian, Plinian, and Strombolian eruptions. HoloVulcano has been developed for Microsoft HoloLens AR device. Wearing the HoloLens, users can interact with the volcano through voice, gazing, and gestures and view different eruptions from different points in the island. HoloVulcano will be used for training emergency exercises and public education.

Keywords: HoloVulcano, Microsoft Hololens, Virtual Reality, Vulcano Island

1. INTRODUCTION & BACKGROUND

Simulations play a significant role in public safety and disaster and emergency training, education, preparedness planning and response management. They provide valuable numeric as well as visual information to users. 3D simulations and animations are one of the most popular ways for natural and technological hazards visualization for non-scientists and general public. Virtual reality (VR), augmented reality (AR) and mixed reality (MR) are the most recent trends in visualization technologies. While AR applications allow users to see the real world with a layer of digital content, MR simulations help users to see the real, physical world and objects but also see and interact with responsive virtual objects (Zhou *et al.* 2008). AR and MR are very important innovative concepts and applications because: 1) they provide powerful, contextual, and situated learning for education and training; 2) they aid exploration of the complex interconnections seen in information in the real world (Zhu *et al.* 2014); and 3) they provide rich context aware surroundings and computing environment for decision making and effective teamwork (Frenk et al. 2010; Thomas et al. 2010; Kalantari and Rauscbnabel 2016).

AR and MR are among the key emerging technologies for education and training over the next few years. They allow "learners to visualize complex spatial relationships and abstract concepts, experience phenomena that is not possible in the real world, interact with two and three dimensional synthetic objects in the mixed reality, and develop important practices that cannot be developed and enacted in other technology-enhanced learning environments" (Mekni and Lemieux 2017: 209).

AR and MR have demonstrated successful applications and specialized niches in a number of sectors, including manufacturing and construction, transportation and entertainment, education, public health and medicine, defense, and public safety and security (Haklay 2001; Carmigniani and Furht 2011; Bacca *et al*. 2014; Stone *et al.* 2017). Public safety and disaster and emergency management is emerging as new and growing application areas of AR and MR technologies (Nunes 2017; Nurmi and Tarkoma 2017; Vassell *et al.* 2017). AR and MR are used in public safety and emergency management mainly to 1) enhance situational awareness; 2) plan for interventions; 3) enhance emergency responders' ability to navigate in degraded visual conditions, 4) train large-scale emergency scenarios, and 5) public education and training.

Increase in the power and decrease in the size of computers has created a new era for development of wearable and mobile AR and MR enabled devices (Brohm *et al.* 2017). As the AR and MR technologies are becoming mature and well established, more advanced and popular applications are developed (Hung-Lin *et al*. 2013). Moreover, their usability and social acceptance are increasing and their limitations are reduced.

Recent studies and experiments show that, for example, 3D animations that reconstruct volcanic processes make scientific results more understandable to wider audiences. End users, such as students, find 3D animation or AR simulations much more interesting and more efficient in knowledge transfer. Moreover, experience show that, scientific knowledge transfer can be enhanced using modern visualization techniques

such as VR and AR, which are more attractive to the new generations (Baldassi *et al*. 2018).
Volcanos are the source for disaster events in many countries, particularly those located around the ring of fire. Considerable number of world population is living in proximity to volcanic hazards (Hlodversdottir *et al.* 2018). Studies show that about 47 million people are living within 5 km of active volcanos. This number increases to 58 million within 10 km, 200 million within 30 km radius and more than 750 million within 100 km of about 1378 active volcanos. Indonesia has the largest population within 100 km of an active volcano (179 million), followed by the Philippines and Japan (approximately 100 million each, based on the 2010 population data) (Cottrell 2014). Annually a number of major volcanic eruptions occur that require mass evacuation, followed by significant human, economic and environmental impacts. According to the global disaster database -EM-DAT, from 1900 to 2018, a total of 245 volcanic disasters have been reported worldwide with 96,306 fatalities, 11,760 injures, and 6,408,600 affected (https://www.emdat.be/).
The remaining sections of this paper are organized as follows: Section two provides background information about the Vulcano Island in Italy covering its volcanic history, hazards, vulnerability and the rational for the HoloVulcano application. Section three will provide details of the application development process which includes methodology and outcome. Section explains some of the use cases for the application. Finally section five concludes the paper with some discussions about the future directions.

2. VULCANO ISLAND

Vulcano Island is a small island in Sisley, Italy, with an area of about 23 km^2. The island is composed of four main built up areas of Vulcanello, Porto, Lentia, and Piano(Figure 1). Vulcanello contains second homes that are occupied on the weekends and the summer mainly. Porto is a mix of commercial, tourism, hotel, and residential activities. Piano is the residential place for majority of permanent residents. Vulcano has a Carabinieri, a diesel and a solar power generation stations, two helipad facilities, one main port (Porto Lavante) and two small ports (Porto Ponente and Jelso). Rapid growth of the island in recent years has increased the risk and highlighted the importance of emergency management and public education.
Vulcano Island is made up of volcanic rocks. It is a truncated cone that was created about 100,000 years ago. There are four active volcanoes on the island, three belonging to the Vulcanello region of the island, and the large La Fossa Cone. The La Fossa crater is located in the center of the caldera with the height of 390m and a base diameter of 1 km (Dongarrà and Varrica 1998). Vulcano is known to have erupted very violently in February 1444.
At that time there was a chain of three islands: Vulcano, Vulcanello and Lipari. Vulcanello was so close to Vulcano that the two eventually became fused together following a subsequent eruption in 1550 (Cheshire 2018).
Since the last eruption of the La Fossa was in 1890, the existing population has no direct experience of the volcanic eruption and their associated potential hazards. In the absent of public education and outreach activities, local residents and visitors do not have sufficient comprehension of the extent, shape, and potential consequences of an eruption. Therefore, communicating the hazard and risk in a way that people can better appreciate the potential risks would be very useful.

Figure 1. Vulcano Island, source: Dellino et al., 2011.

During the last 1000 years, La Fossa volcano experienced two sub-Plinian eruptions with similar intensities and magnitudes and at least 8 long–lasting Vulcanian cycles (Biass *et al*. 2016). The last eruption of the La Fossa occurred between 1888 and 1890 (Figure 2). The eruption consisted of pyroclastic material and bread-crust bombs, which normally describes the Vulcanian type of eruptions. Studies show that future eruptions are possible (Dellino *et al*. 2011).


Figure 2 Eruption of the La Fossa in Vulcano 1888-1890. (Mercalli and Silvestri 1891)

3. ERUPTION SCENARIOS AND VOLCANIC HAZARDS

3.1. Eruption Scenarios

HoloVulcano aims to simulate several types of eruptions and states including normal state, unrest state, single Vulcanian, long lasting Vulcanian, Plinian, and

Strombolian eruptions. Normal state is the current state of the La Fossa where intensive fumerals are observed near the La Fossa Cone (Figure 3) and in some areas near the Porto.

Figure 3 Normal activities of the La Fossa (source: author)

Unrest scenario corresponds to a higher than normal fumerals, seismic activities and increasing temperature. Unrest situation may be for a short period (few days) or a long period (several months). Unrest situation or scenario may or may not lead to a larger eruption. However, seismic activities and gassing could be harmful to people and properties near the volcano during the unrest phase.

Vulcanian eruption is a short, violent, relatively small explosion of viscous magma. This type of eruption results from the fragmentation and explosion of a plug of lava in a volcanic conduit, or from the rupture of a lava dome. Vulcanian eruptions create powerful explosions in which material can travel faster than 350 meters per second (800 mph) and rise several kilometers into the air. They produce tephra, ash clouds, and pyroclastic density currents. Vulcanian eruptions could be single or repetitive and continue for days, months, or years. The name Vulcanian eruption comes after the Vulcano Island itself (Ball 2018).

Plinian eruption is one of the most violent, impressive, and destructive eruptions due to pyroclastic flow. The term Plinian comes from Pliny the Younger, a Roman nobleman, who witnessed the eruption of Vesuvius in 79 A.D. that destroyed Pompeii and Herculaneum (Biass *et al.* 2016). Plinian eruptions vary in energy, size, and distribution (Dellino *et al*., 2011). They are caused by the fragmentation of gassy magma, and are usually associated with very viscous magmas. They release enormous amounts of energy and create gas columns and mushroom like ash clouds that can rise up to 50 km high at speeds of hundreds of meters per second. Ash from an eruption column can drift or be blown hundreds or thousands of kilometers away from the volcano (Dellino et al., 2011). They can produce falls of ash, scoria and lava bombs kilometers from the volcano, and pyroclastic density currents that raze forests, strip soil from bedrock and obliterate anything in their paths. Researchers believe that Plinian and sub-Plinian eruptions with plume heights between 7 and 8 km are possible for the LaFosa (Bonadonna 2006; Biass *et al*. 2016).

Strombolian eruptions, named after the Stromboli Volcano (Italy), are usually mild explosive eruptions. Strombolian eruptions are distinct bursts of fluid lava from the mouth of a magma-filled summit conduit (Ball 2018). The eruptions may continue with little variation for years. The lava explosions can reach as high as hundreds of meters (Ball 2018).

Volcanic eruptions have different sizes that can be measured by magnitude (the total amount of erupted materials), intensity (volcanic materials rate of eruption), dispersive power (area covered by erupted materials), and destructive potential (Blong 1984). Each eruption can involve a number of primary and secondary hazards.

3.2. Volcanic Hazards

Lava flows, projectiles, tephra falls, pyroclastic flows, debris avalanches, lahars, earthquake and ground deformation, tsunami, acid rain, gases are among the key hazards associated with volcanos (Blong, 1984) (Figure 4). Some of the main hazards, especially those that have been used in HoloVulcano application are described in further details in this section.

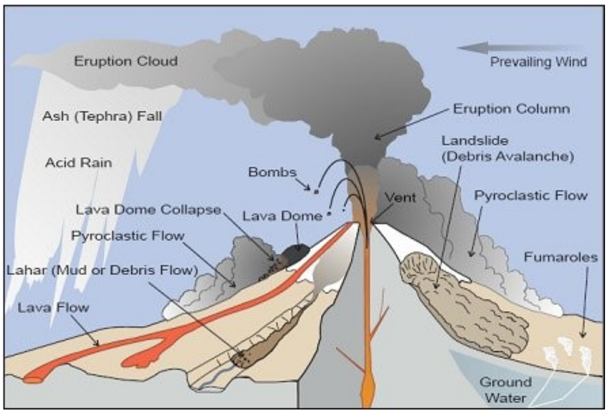

Figure 4 Volcanic hazards (source: http://www.geocoops.com/volcanic-activity1.html)

A volcanic eruption usually generates lots of loose unconsolidated debris. When these materials are mixed with water (rainfall, melting of snow or ice), mudflows or lahars are created. Lava flows are common in Strombolian eruptions. Although they have been known to travel as fast as 64 km per hour, most of them are slower and give people time to evacuate. Therefore, lava flows are most damaging to properties, infrastructure, agriculture, and the environment.

Volcanic bombs, blocks and lapilli which leave the vents with ballistic trajectories are referred to as projectiles. Ballistic projectiles are ejected in various sizes and travel distances. Ballistic projectiles in the shape of bombs or blocks are common in Vulcanian eruptions and are known to have been prominent in the last eruption of La Fossa of 1888-1890 (Biass *et al.*, 2016). Depending on their size and temperature, ballistics could generate different human and property

impacts. Ballistics can also create secondary hazards such as wildfire in densely vegetated areas.
Pyroclastic flows are one of the most dangerous aspects of volcanism. The turbulent flow of heavy particles mixed with gas bubbling, create pyroclastic flows. The temperature ranges between 200 to 700 degrees Celsius. They cause death by suffocation and burning. They can travel very fast so that few people can escape. Pyroclastic flows are common to Vulcanian eruptions and are closely associated with the La Fossa.
Tephra is the most widespread of all volcanic hazards (Biass *et al*. 2016). Tephra cover surfaces like snow and can be very destructive. Tephra deposits are very dense and can cause the collapse of roofs. They can normally affect areas far from the volcano. Tephra can destroy vegetation, and farms, and can kill livestock that eat the ash covered vegetation. Tephra falls can disrupt agriculture for several years after the eruption. Tephra can damage and disrupt water and electricity infrastructure, and civil aviation (Biass *et al*. 2014; Bonadonna *et al*. 2012).

4. APPLICATION DEVELOPMENT

Figure 5 shows the overall process of developing the HoloVulcano application for Microsoft HoloLens.

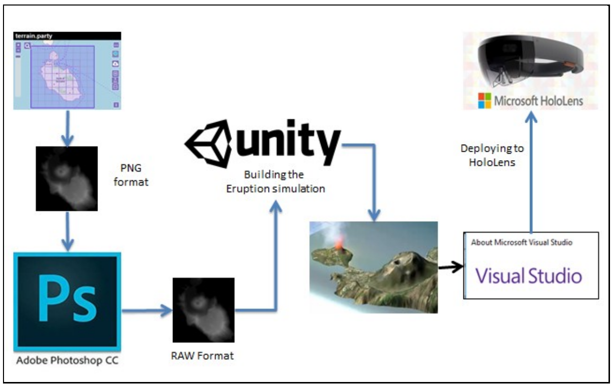

Figure 5 Process of developing HoloVulcano

TerrainParty (www.terrain.party) was used to extract the elevation data for Vulcano Island. Terrain Party allows users to extract elevation data in PNG format that can be used for 3D terrain modeling (Figure 6).

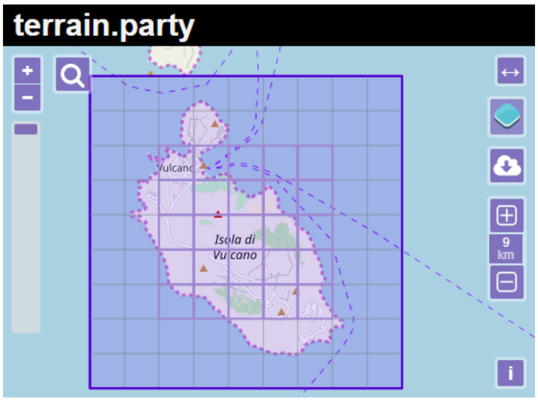

Figure 6 Extracting elevation PNG file from terrain.party

The terrain party output needs to be converted to RAW format to be used for terrain modelling in Unity. Adobe Photoshop was employed to convert the PNG elevation data to RAW format. The RAW elevation file was then imported and added to a terrain object in Unity to make the 3D model of the Island (Figure 7).

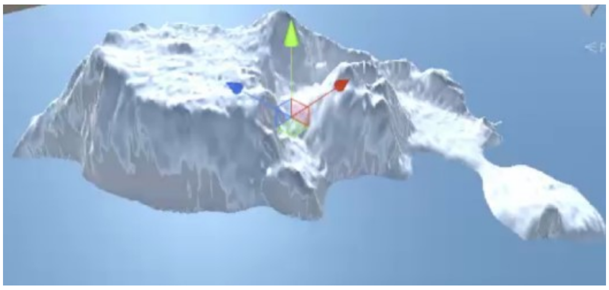

Figure 7 Initial terrain model of the Vulcano Island in unity

The 3D terrain can be textured using existing satellite imagery or texture editing tools. In this version of HoloVulcano a google map satellite image has been used to texture the 3d terrain model (Figure 8).

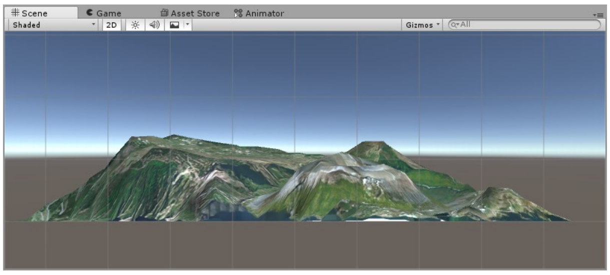

Figure 8 3D model of Vulcano Island developed in Unity

Eruption scenarios were then developed and added to the 3D model using various tools in Unity and third party unity assets that were imported to Unity. For the most parts Unity's particle system was used to simulate different eruption scenarios.
For example, Figure 9 shows the simulation of normal conditions at La Fossa. Unity particle system was used to simulate the normal activities (fumerals).

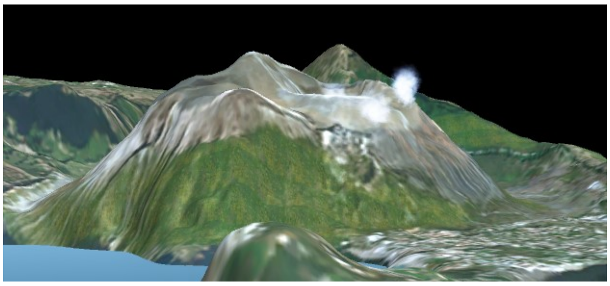

Figure 9 Simulation of normal activities of La Fossa in Vulcano Island

The unrest situation is an extension of normal condition with higher emission rates of volcanic gases and fumaroles (Figure 10).

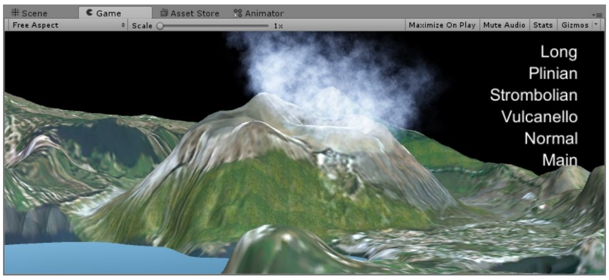

Figure 10 Simulation of Unrest activities of the La Fossa

Snapshots of the Strombolian and Plinian eruptions are shown in Figures 11 and 12. They include simulations of lava fountain, lava flow, pyroclastic flow, and ballistic projectiles.

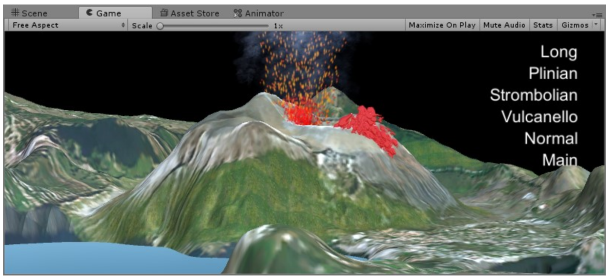

Figure 11 Simulation of Strombolian eruption of La Fossa

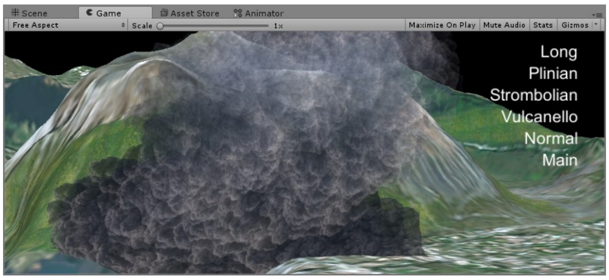

Figure 12 Simulation of Plinian eruption with pyroclastic flow

Users can navigate in the application and interact with the HoloVulcano using voice and air tap. A menu helps the user to select an eruption scenario and choose the desired viewing location (from north, east, south, west, top, etc.). Users can also use the voice command to move backward, forward, left and right.
The simulation application built in the Unity environment is then opened in Microsoft Visual Studio to create the Hololens version of the application. HoloLens is the first fully self-contained, lightweight portable wearable system, enabling users to interact with high definition holograms that are very much different from the existing augmented and mixed reality technologies (Figure 13). Users can wear the 580-gram HoloLens headset that runs Windows 10 to display virtual or holographic objects (Asgary 2017). HoloLens results are the best mixed-reality experience to date by a standalone, untethered device (Furlan 2016). HoloLens comes with an Intel Atom processor and 3D spatialized sound, Wi-Fi connectivity, a depth and color camera. It offers a 120 degree spatial sensing system and a see-through screen for each eye (Orlandi *et al*., 2018). Hololens also features a Holographic Processing Unit (HPU) that combines real-world and holographic data. Compared to other AR devices, HoloLens demonstrates enhanced integration capabilities (Orlandi *et al*., 2018).

Figure 13 Microsoft Hololens kit

5. HOLOVULCANO USE CASES
HoloVulcan application can be used in disaster and emergency training and exercises, public education and awareness razing, and hazard tourism.

5.1. Emergency Exercise and Training
Like many other fields, training of the current and future emergency managers is one of the main applications of the AR technologies. As discussed earlier, simulation-based learning is very effective in disaster and emergency management. It promotes development of experience without having an actual disaster situation. Learners can be trained without any risk to them and others. AR simulations can be enhanced to provide details of emergency scenarios with virtual human and economic losses. Moreover, trainees will benefit from a virtual 3D view of the scene, which will help them with a better spatial understanding of the situations.
Emergency exercises such as table top exercises are key elements of emergency preparedness planning. Functional and table top exercises, whether for training of future emergency managers or for emergency preparedness planning can benefit from AR and MR visualization technologies. These tools enable exercise participants to have a common understanding of the emergency scenario. As such, one of the key use cases of the HoloVulcano application is for emergency training and exercise. In particular, this application will be used for CERG-C (CERG 2011) field training program that is held every year in Vulcano Island. During this field trip participants conduct hazard, vulnerability and risk analyses and take part in an emergency table top exercise (Figure 14, 15).

Figure 14 Table top exercises with CERG-C participants in Vulcano Island, May 2018.

Figure 15 Demonstration of HoloVulcano in table top exercise with CERG-C participants May 2018

5.2. Public Education and Awareness Raising

AR and MR technologies are expected to have increasing applications in education. The more interactive the applications are the more interesting they will be for learners. These technologies facilitate learning through entertainment, deep-understanding of the complex concepts, interaction with virtual objects, and real-time feedback (Ashby 2014). AR and MR simulations enable teachers to explain complex and significant scientific information (such as volcanic processes and eruptions) to students in a much easier and attractive and safer methods. Moreover, applications can be tailored to meet the needs of different types of students with different learning and physical abilities (Alzahrani 2018).

During the annual CERG-C field trip in Vulcano a public education activity is conducted with the school children (Figure 15). This is a role playing exercise conducted in collaboration with the Italian Civil Protection with the goal of awareness raising among the local population on topics such as volcanic hazards and risk, emergency preparedness and management, and the importance of collaboration between citizens and official institutions, such as the Civil Protection (Bonadona, *et al*. 2014). HoloVulcano can be used in the future public education activities such as this to enhance the exercise and participants experience.

Figure 15 CERG-C public education activities in Vulcano's elementary school

5.3. Volcanic Tourism

AR and MR can have huge applications in tourism industry. Already, several tourism related HoloLens applications, particularly for historic and cultural sites have been developed (Debandi 2018). These solutions are used to enhance tourists' experiences of places that they visit. Volcanoes are among the most popular tourist destinations. Studies show that annually more than 130 million tourists visit nine major volcanos including Fuji-Hakone in Japan, Mt. Teide National Park in Spain, Yellowstone National Park in the United States, Tongariro in New Zealand, and Vesuvius in Italy (Erufurt-Cooper 2011). Volcanos attract tourists because of their spectacular views and adventure activities (D'Alessandro *et al*. 2013). Research also shows that increased volcanic activity draws more tourists to volcanic sites (Erfurt-Cooper 2014).

While very interesting and impressive, there are health and safety risks associated with volcano tourism (Heggie 2009). Volcanic hazards such as gases, tephra, landslides, lahars, and lava flows can harm tourists. As such, it is very important that tourists are educated on health and safety issues associated with volcanos. Vulcano has integrated different volcanic tourist elements to create a unique destination to entice visitors. The La Fossa cone is one of the key attractions for tourists visiting the Vulcano Island and tourism is Vulcano's main industry (Biass *et al*. 2016).

HoloVulcano application can be used by tourists visiting the island in order to gain a better understanding of the volcanos behavior. For example, hotels and shops can provide AR devices with HoloVulcano application to their clients.

6. CONCLUSIONS AND FUTURE WORKS

HoloLens is an AR tool with significant potentials for disaster and emergency and public safety applications. Visualizing disaster and emergency simulation scenarios in AR format presents a new level of realism. As more and more low cost and compact wearable AR hardware are becoming available, new and interesting applications are developed and used in public safety and emergency management field.

HoloVulcano is an AR application which simulates various possible eruption scenarios of the La Fossa volcano for emergency management training, public education, and tourism. While the effectiveness of this application has not been fully tested yet, it is believed that the application can provide a new way of visualizing volcano eruptions and emergency scenarios that can enhance learning and training.

More features will be added to the current application to make it more scientific, interactive, and educational. For example, additional features will be added that enable users to set different values for different parameters of each eruption scenario.

AUTHORS BIOGRAPHY

Dr. Ali Asgary is an Associate Professor and currently the Associate/Executive Director of York University's Advanced Disaster, Emergency and Rapid-response Simulation (ADERSIM). He is a co-founder of disaster and emergency management programs in York University and Brandon University in Canada. He has served as the area coordinator and the Graduate Program director of the Disaster and Emergency Management at York. He is the Co-PI for York University's ADERSIM program. Dr. Asgary is an expert in disaster, emergency, and business continuity management with focus on simulations and exercises. His extensive research and teaching are enhanced by his active contributions to the profession and by translating them into the real world practices at different levels. His research has been funded by ORF (Ontario Research Fund), NSERC, NSERC CREATE, SSHRC, GEOIDE, PreCarN, AIF (Academic Innovation Fund), etc. He is the author or co-author of numerous scholarly and practitioner articles in various aspects of disaster and emergency management particularly in disaster and emergency evacuation simulation at small, medium and large scale environments such as buildings, university campuses/schools, islands, airports, subway stations, and cities. Dr. Asgary has received different awards for his research, teaching and other contributions, including the International Association of Emergency Management Award and the outstanding paper of the year award by the Journal of Disaster Prevention and Management. Dr Asgary is frequently consulted by media (examples include: CBC, Toronto Star, Gloabe & Mail, CTV News, CBC Radio, Vancuver Sun, etc.), associations, governmental and non-governmental organizations and private sector companies – for the latest developments and trends in disaster events, disaster management and business continuity in Canada and globally. Dr. Asgary served as a member of the Global Board of the International Association of Emergency Managers and the President of the Canadian Council of the International Association of Emergency Managers (IAEM-Canada) from 2007 to 2009.

REFERENCES

Alzahrani, O., 2018. Evaluation of the effectiveness of using mixed reality technology in inclusive education. IEEE (http://www.cs.bham.ac.uk/~rjh/courses/ResearchTopicsInHCI/2017-18/Coursework/alzahraniohoud.pdf)

Behzadan, A. H., and Kamat, V.R., 2017. Enabling discovery-based learning in construction using telepresent augmented reality. Automation in Construction, 33, 3-10.

Asgary A., 2017. Holodisaster: Leveraging Microsoft Hololens in Disaster and Emergency Management. IAEM Bulletin, January 2017.

Ashby, C., Burns, J., and Royle, J., 2014. All kids can be readers: The marriage of reading first and inclusive education. Theory Into Practice, 53 (2), 98-105.

Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk, 2014. Augmented reality trends in education: A systematic review of research and applications. Journal of Educational Technology & Society, 17(4), 133-149.

Baldassi, S., Kohno, T., Roesner, F., and Tian, M., 2018. Challenges and New Directions in Augmented Reality, Computer Security, and Neuroscience -- Part 1: Risks to Sensation and Perception, eprint arXiv:1806.10557.

Ball, J., 2018. Types of Volcanos. Geology.com, (https://geology.com/volcanoes/types-of-volcanic-eruptions/).

Biass, S, Scaini C, Bonadonna C, Folch A, Smith K, Höskuldsson A., 2014. A multi-scale risk assessment for tephra fallout and airborne concentration from multiple Icelandic volcanoes - Part 1: Hazard assessment. Nat Hazards Earth Syst Sci., 14(8), 2265–287.

Biass, S., Bonadonna, C., Connor, L., and Connor, C., 2016. TephraProb: a Matlab package for probabilistic hazard assessments of tephra fallout. Journal of Applied Volcanology Society, 5(10).

Blong, R.J., 1984. Volcanic Hazards: A Sourcebook on the Effects of Eruptions provides a comprehensive discussion of volcanic eruptions and their effects. Academic Press.

Bonadonna, C, Folch A, Loughlin S, Puempel H., 2012. Future developments in modelling and monitoring of volcanic ash clouds: outcomes from the first IAVCEI-WMO workshop on Ash Dispersal Forecast and Civil Aviation. Bull Volcanol. 74(1), 1–10.

Bonadonna, C., 2006. Probabilistic modelling of tephra dispersion. In: Mader, HM, Coles SG, Connor CB, Connor LJ, editors. Stat Volcanol. London: Geological Society of London: 2006. p. 243–59.

Bonadonna, C.; Consuegra, D.; Duvernay, B.; Fäh, D.; Frischknecht, C.; Gregg, C.; Jaboyedoff, M.; Lateltin, O.; Menoni, S.; Franco, R.; Rosi, M., 2014. The Cerg-C: A Specialisation Certificate in Geological and Climate Related Risk of the University of Geneva, Switzerland, American Geophysical Union, Fall Meeting 2014, abstract id. NH51A-3842.

Brohm D., Domurath N., Glanz-Chanos, V., Grunert, K.G, 2017. Future trends of augmented reality, in Augmented reality for food marketers and

consumers edited by Leanne W.S. Loijens, Wageningen Academic Publishers.
Carmigniani J, Furht B., 2011, Augmented reality: an overview. In: Furht B, editor. Handbook of augmented reality. New York: Springer; 2011. pp. 3–46.
CERG, 2011. Training on Risk Management: The Case Study of Vulcano Island, Italy. Risk Workshop, Geneva (https://www.unige.ch/sciences/terre/CERG-C/files/8314/7732/2217/CGREGG_CERGC2011.pdf).
Cheshire, G.E., 2018. Linguistically dating and locating the origin of Manuscript MS408. Science Survey, 2, 1-36.
D'Alessandro, W.D, Aiuppa, A., Bellomo, S., Brusca, L., Clabrese, S. Kyriakopoulos, K…Longo, M. , 2013. Sulphur-gas concentrations in volcanic and geothermal areas in Italy and Greece: Characterising potential human exposures and risks. Journal of Geochemica Exploration, 131, 1-13.
Debandi, F., Iacoviello, R., Messina, A., Montagnuolo, M., Manuri, F., Sanna, A., and Zappia, D., 2018. Enhancing cultural tourism by a mixed reality application for outdoor navigation and information browsing using immersive devices, IOP Conference Series: Materials Science and Engineering, 364 (1).
Dellino, P., De Astis, G., La Volpe, L., Mele, D., Sulpizio, R., 2011. Quantitative hazard assessment of phreatomagmatic eruptions at Vulcano (Aeolian Islands, Southern Italy) as obtained by combining stratigraphy, event statistics and physical modelling. Journal of Volcanology and Geothermal Research, 201(1–4), 364-384.
Dongarrà, G., and Varrica, D., 1998, The presence of heavy metals in air particulate at Vulcano island (Italy), Science of The Total Environment, 212(1), 1-9.
Erfurt-Cooper, P. (2011). Geotourism in volcanic and geothermal environments: Playing with fire? Geoheritage, 3, 187–193.
Erfurt-Cooper, P., 2014. Introduction. In P. Erfurt-Cooper (Ed.), Volcanic Tourist Destinations (pp. 3-34) Heidelberg, Germany: Springer.
Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, Fineberg H, Garcia P, Ke Y, Kelley P, Kistnasamy B, Meleis A, Naylor D, Pablos-Mendez A, Reddy S, Scrimshaw S, Sepulveda J, Serwadda D, Zurayk H., 2010. Health professionals for a new: transforming education to strengthen health systems in an interdependent world. Lancet, 376(9756), 1923–1958.
Furlan, Rud, 2016. The Future of Augmented Reality: Hololens Microsoft's AR headset shines despite rough edges. SPECTRUM.IEEE.ORG, North American, June 2016, page 21.
Gianni, M., Gonnelli, G., Sinha, A., Menna M., and Pirri, F., 2013. An Augmented Reality approach for trajectory planning and control of tracked vehicles in rescue environments,", *2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Linkoping, 2013, 1-6.
Haklay ME., 2001. Virtual reality and GIS: applications, trends and directions. In: Fisher P, Unwin D, editors. Virtual reality in geography. London: CRC Press; 47–57.
Haynes, P.S., and Lange, E., 2016. In-situ flood visualisation using mobile AR,. *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, Greenville, SC, 2016, 243-244.
Heggie, T.W., 2009. Geotourism and volcanoes: Health hazards facing tourists at volcanic and geothermal destinations. Travel Medicine and Infectious Disease, 7, 257-261.
Hlodversdottir, H., Thorsteinsdottir, H., Bjork, E.T., Njardvik, U., Petursdottir, G., And Hauksdottir, A., 2018. Long-term health of children following the Eyjafjallajökull volcanic eruption: a prospective cohort study, European Journal of Psychotraumatology, 9:sup2.
Hung-Lin Chi, Shih-Chung Kang, Xiangyu Wang, 2013. Research trends and opportunities of augmented reality applications in architecture, engineering, and construction, Automation in Construction, 33, 116-122.
Irizarry, J., Gheisari, M., Williams, G., and Walker, B.N., 2013. InfoSPOT: A mobile Augmented Reality method for accessing building information through a situation awareness approach, Automation in Construction, 33, 11-23.
Kalantari M., Rauschnabel P., 2018. Exploring the Early Adopters of Augmented Reality Smart Glasses: The Case of Microsoft HoloLens. In: Jung T., tom Dieck M. (eds) Augmented Reality and Virtual Reality. Progress in IS. Springer, Cham
Karbovskii, V., Voloshin, D., Karsakov, A., Bezgodov, A., Gershenson, C., 2016. Multimodel agent-based simulation environment for mass-gatherings and pedestrian dynamics. Future Generation Computer Systems, 2016.
Kaufmann,H., 2003. Collaborative Augmented Reality in Education, Position paper for keynote speech at Imagina 2003 conference, Monaco (2003).
Kim, W; Kerle, N; Gerke, M., 2016. Natural Hazards and Earth System Sciences; Katlenburg-Lindau 16(1), 287-298.
Mekni M, and Lemieux, A., 2014. Augmented Reality: Applications, Challenges and Future Trends, University of Minnesota, Crookston Campus, Applied Computational Science. http://www.wseas.us/e-library/conferences/2014/Malaysia/ACACOS/ACACOS-29.pdf
Mercalli, G., Silvestri, O., 1891. L'eruzione dell'Isola di Vulcano incominciata il 3 agosto 1888 e terminata il 22 marzo 1890. Annali dell'ufficio Centrale di Meteorologia e Geodinamica, 10, 71–281.

Mitsuhara, H.; Iguchi, K.; Shishibori, M., 2017. Using Digital Game, Augmented Reality, and Head Mounted Displays for Immediate-Action Commander Training. International Journal of Emerging Technologies in Learning . 12(2), 101-117.

Nevatia, Y., *et al*., 2008. Augmented autonomy: Improving human-robot team performance in Urban search and rescue. *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, 2008, 2103-2108.

Nunes, I.L., Lucas, R., Simões-Marques, M., and Correia, N., 2017. Augmented Reality in Support of Disaster Response. International Conference on Applied Human Factors and Ergonomics AHFE 2017, Advances in Human Factors and Systems Interaction, 155-167.

Nurmi, P.; Tarkoma, S., 2017. Low-cost support for search and rescue operations using off-the-shelf sensor technologies. Electronics Letters, 53, (15), 1011-1013.

Orlandi, L., Seyegnani D., and Conci, N., 2018. Improving the efficiency of on-site operators in utility management: combining HoloLens and AR for real-time check of electricity meters. Electronic Imaging, Visual Information Processing and Communication IX, 264-1-264-5.

Owaidah, Almoaid A., 2015. Hajj crowd management via a mobile augmented reality application: a case of The Hajj event, Saudi Arabia. MSc(R) thesis, University of Glasgow.

Stone, R.J., Guest R., Mahoney, P., Lamb, D., and Gibson, C., 2017, A 'mixed reality' simulator concept for future Medical Emergency Response Team training, BMJ Publishing Group Limited.

Suyang Dong, Chen Feng, Vineet R. Kamat, 2013. Sensitivity analysis of augmented reality-assisted building damage reconnaissance using virtual prototyping, Automation in Construction, 33, 24-36.

Thomas RG, John NW, Delieu JM. 2010. Augmented reality for anatomical education. Journal of Visual Communication in Medicine, 33(1), 6–15.

Vassell, M., Apperson, O., Calyam, P., Gillis, J., and Ahmad, S., 2016, Intelligent Dashboard for augmented reality based incident command response co-ordination, *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, 2016, pp. 976-979.

Zhou F, Duh HB, Billinghurst M., 2008. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. 7th IEEEACM International Symposium on Mixed and Augmented Reality. 2008;2(4), 193–202.

Zhu E, Hadadgar A, Masiello I, Zary N., 2014. Augmented reality in healthcare education: an integrative review. PeerJ2:e469; DOI10.7717/peerj.469

Zhu E, Hadadgar A, Masiello I, Zary N., 2014. Augmented reality in healthcare education: an integrative review. *PeerJ* 2:e469

ISBN 978-88-85741-12-6; Bruzzone and Sottilare Eds.

INTELLIGENT TUTORING IN THE WILD: LEVERAGING MOBILE APP TECHNOLOGY TO GUIDE LIVE TRAINING

Benjamin Goldberg[a], Nicholas Roberts[b], W. Gabe Powell[c], and Elyse Burmester[d]

[a]Natick Soldier Research and Development Center – Simulation and Training Technology Center
[b,d]Dignitas Technologies
[c]United States Military Academy

[a]benjamin.s.goldberg.civ@mail.mil; [b]nroberts@dignitastechnologies.com; [c]william.powell@usma.edu; [d]eburmester@dignitastechnologies.com

ABSTRACT
Mobile computing technologies are extending how people can interact with educational and training content in a whole new way. Through high resolution displays, intuitive user interfacing, embedded sensing technologies, and well supported app development communities, there is a plethora of content that can be used to build effective materials that target knowledge and skill development. To truly enhance this new training paradigm, extending Intelligent Tutoring System (ITS) to support mobile interactions can provide a new means to managing training in a rich contextualized environment. In this instance, learning takes place in the natural environment where directed experiences focus on the elements surrounding one's location. In this paper, we describe the development of a new mobile ITS application using the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Goldberg, Brawner & Holden, 2012). The domain of land navigation was applied as a use case, with direct support for the United States Military Academy at West Point. We describe the training concept, how GIFT was extended to support this concept from an architectural and assessment standpoint, along with implementation plans for an initial training effectiveness study.

Keywords: workstation design, work measurement, ergonomics, decision support system

1. INTRODUCTION
Intelligent Tutoring Systems (ITS) are computer-based applications that apply artificial intelligence modeling techniques to provide individualized learning experiences. They are traditionally developed for desktop and laptop computing environments, with empirical evaluations showing effect sizes up to 1 standard deviation in well-defined academic domains (e.g., physics, algebra, computer programming, etc.) when compared against traditional instructional methods (VanLehn, 2011). With investments across the Department of Defense to leverage ITS methods to enhance Simulation-Based Training (SBT) methods, it is recognized that new interaction modalities are required to compensate for the interaction space across military training contexts.

As interfacing technologies continue to mature in mobile contexts, smartphone technology is prime for extending ITS benefits into less traditional delivery formats. This new paradigm also provides a new context by which a learner can interact with physical environments beyond the traditional classroom.

In this paper, we present work associated with the development of an ITS mobile application (app) that trains land navigation skill sets. The app is designed around the GIFT specification (Sottilare, Goldberg, Brawner & Holden, 2012), and was implemented in a generalizable fashion that can extend across many domains. We begin by describing the vision behind this work from a learning science perspective, followed by modifications to the GIFT architecture to support the implementation. We then conclude with an overview of how the app was applied for a land navigation use case, along with the study design used to gauge its impact.

2. THEORY BEHIND SIMULATION AND TRAINING
The benefit associated with SBT platforms is they provide realistic environments that allow individuals to master complex material and learn and apply new information through execution of simulated tasks (Menaker, Coleman, Collins, & Murawski, 2006). The learning process is influenced by student-centered teaching methods prompted by theories of 'discovery' (Bruner, 1966; Hermann, 1969) and 'active' (Johnson, Johnson, & Smith, 1991) learning. They incorporate interacting elements of logic, memory, visualization, and problem solving that cater to elements required for learning; engagement, interaction, and satisfaction (Amory, Naicker, Vincent, & Adams, 1999). This is achieved by replacing traditional classroom instructional techniques with methods of role-playing, simulations, self-regulated exercises, and other types of problems requiring creative and critical thinking skills (Greitzer, Kuchar, & Huston, 2007). Research has

demonstrated these forms of training to be an effective alternative to traditional classroom instruction because they assist learners in rapidly creating and adjusting mental models for newly acquired information (Cuevas, Fiore, Bowers, & Salas, 2004). These environments also provide a forum for learners to actively participate with learning material and to view the effect varying actions have on outcomes.

2.1. Learning by Doing

The application of SBT across military domains allows for the development of authentic scenarios that facilitate learning and cognitive development. In the military context, interactions in a simulated exercise enable visualization and practice of task execution. As a result, learners come to their first live performance experience with an advantage (Waldman, 2009). In addition, SBT enables Soldiers to interact with multiple scenarios in a short timeframe. This allows for rapid exposure to variations in task conditions that build task relevant experience (Pine, 2009). In these instances, executing live simulated exercises is ideal; however, they are traditionally confined to indoor environments that use virtual interfacing to create the stimuli to interact within.

While technology-enabled SBT has seen success in virtual and constructive application environments, applying SBT technologies to facilitate self-regulated live training exercises has not received extensive attention. Yet, with advancements in adaptive instructional systems, and the maturity level of the Generalized Intelligent Framework for Tutoring (GIFT), adaptive training pedagogical practices can now be embedded in smartphone apps using cloud-based infrastructures.

A primary goal of many modern military training systems is to provide the learner with strategies that aid in the development of higher-order thinking skills and enable them to adapt decision-making tactics under variable missions and conditions (Wisher, Macpherson, Abramson, Thorton, & Dees, 2001). In today's combat environment, tasks are executed under a multitude of complex, stressful, and ambiguous settings where decisions must be quick and actions must be executed in a timely manner (Salas, Priest, Wilson, & Burke, 2006). Therefore, training aims to foster successful task execution and the values associated with making reasonable decisions under difficult circumstances (Bratt, 2009). SBT fosters this type of learning by applying principles of instructional design through the processes of development, application, and evaluation of task relevant KSAs in realistic situations (Oser, Cannon-Bowers, Salas, & Dwyer, 1999).

SBT simply replicates a real-world representation of a problem space where knowledge, skills, and abilities can be applied within bounded realistic conditions that aid in skills training (e.g. time pressure, stress). However, simulating a task in a physical environment and providing the ability for an individual to practice does not on its own increase expertise (Ericsson & Ward, 2007). Expertise development in SBT platforms is not practical without apt pedagogical support (Ericsson & Ward, 2007). This is a recognized gap because too often simulations are fielded without pedagogical components and functions. Simulations intended for education and training provide a means for practicing KSAs, but as mentioned above often lack elements of pedagogy that guide the learning process (Nicholson, Fidopiastis, Davis, Schmorrow, & Stanney, 2007). This limitation is the forcing function for advancements in ITS applications within military domain spaces. To meet this evolving need, ITS applications are required for task spaces that involve a combination of cognitive and psychomotor components, where tasks are executed in dynamic environments. To meet this training need, extending ITS functions to mobile platforms is critical.

2.2. A Use Case: Land Navigation

Land navigation is an excellent use case to inform initial development efforts of a mobile adaptive instructional system. This is due to the nature of domain's knowledge components and how current training programs are executed. The initial objective is to replicate exercises and scenarios that are currently instructor led and utilize elements of the natural environment. As a grounding function, we partnered with the United States Military Academy's (USMA) Department of Military Instruction (DMI) to determine design requirements. DMI is responsible for establishing training programs administered to new cadets that focus on foundational skills required for functioning within the Army. For land navigation, a five-day program is administered that incorporates classroom instruction, virtual SBT in a game-based environment, live instructor-led training outdoors, and live practice opportunities all before a culminating qualification event.

A specific event administered on day two of training is what is called a Terrain Walk. The established method of a Terrain Walk incorporates one instructor to a group of learners; in this case a squad of new cadets. The new cadets are guided up a mapped out path with training interventions initiated at specific points along the trail. At each point, a new training objective is targeted, where the instructor engages the group through conversational cueing. The goal of the exercise is to reinforce material covered in classroom instruction through contextualized active learning.

One utility an ITS in this domain aims to provide is self-regulated interactive scenarios that promotes skill acquisition for each individual trainee. This involves providing scaffolds for learners who do not possess the skills to conduct the tasks, and challenging those with prior experience. While the current collective nature of the training allows for direct learner interaction with subject matter experts with years of experience, it reduces hands-on opportunities for all members of a squad, and potentially results in disengaged learners. To address this, the app is being designed to replicate tasks performed in the traditional form, but providing a mechanism that allows each individual learner to perform each scenario at their own pace. In the following sections we review the developmental efforts applied to create the first iteration of a Terrain Walk oriented mobile ITS app.

3. BUILDING A LAND NAVIGATION APP

3.1. Generalized Intelligent Framework for Tutoring (GIFT)

GIFT is a modular framework of software tools designed to simplify authoring ITS applications by providing a configurable tutoring engine that can be integrated with other software (see Figure 1). GIFT has been used in the past to perform assessments based on learners' movements in simulated 3D environments and provide instructional feedback based on their actions within said environments. This same principle of tracking learners' movements in simulated environments can be applied to interactions with mobile devices in live environments. Instead of monitoring learners' positions in a simulated setting, GIFT can take advantage of mobile devices' built-in Global Positioning System (GPS) services to track learners' real-world locations and perform the same assessment operations that it would in a simulated environment. This is the goal that was used to guide the changes made to the GIFT architecture and the development of the mobile ITS application.

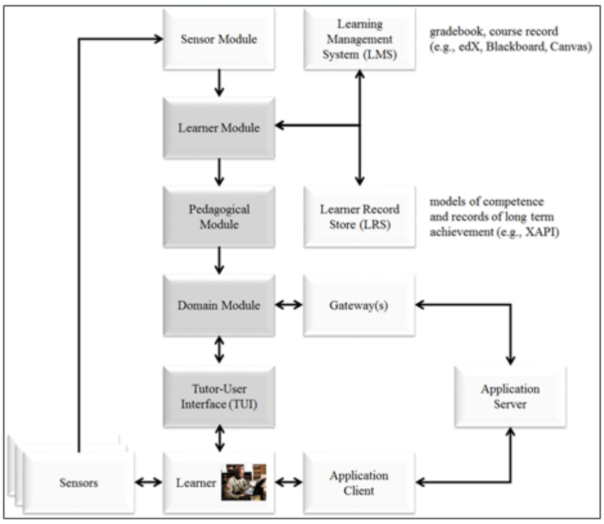

Figure 1: The GIFT Framework Represented by its Modular Components and Their Message Flow Dependencies

3.2. Architecture Modifications

While GIFT has been used before to track the locations of learner entities in simulated environments, the geolocation data passed from mobile devices uses a geodetic (GDC) coordinate system that GIFT's assessment logic was not originally designed to handle. To accommodate this new coordinate system, GIFT's internal messaging logic had to be modified to properly encode and decode the coordinates received from mobile devices' GPS services.

Mobile devices also provide features to display notifications and vibrate that GIFT can take advantage of to provide additional hands-free feedback to learners; features that are particularly useful for conducting live training since learners may not always be looking at the mobile device's screen. To access these features, GIFT's internal messaging system was further modified in order to add new command messages that tell the mobile device to display notifications and vibrate the phone. These new messages were then incorporated into the existing logic that GIFT uses to display messages in other software applications, since this same messaging pipeline would eventually be used to build the gateway needed to communicate with the phone's operating system and gather GPS data from cellular networks

Additional changes were also made to GIFT's authoring tools and domain module. A new type of course object was added to GIFT's course creator to allow authors to use the new mobile ITS app as an inherent training environment interacting with GIFT. Unlike most training applications, GIFT itself was not planned to run alongside the app on a mobile device's operating system, since GIFT cannot currently be installed on mobile devices.

Instead, an instance of GIFT was launched on a publicly-facing server so that the mobile ITS app could display that GIFT tutoring interface as an embedded web page. This approach allows learners to take courses with GIFT and interact with the mobile ITS app at the same time, without installing any GIFT software to their mobile device itself. The approach also posed a challenge for the logic GIFT normally uses to communicate with training applications, since a training application would typically use an instance of GIFT's gateway module running on the same operating system to communicate with the rest of GIFT's modules. Fortunately, work from a previous effort had set up a communication mechanism in GIFT's tutoring interface to allow web-based Unity applications to communicate with GIFT without using the gateway module, effectively using GIFT's tutor module as a surrogate. This communication mechanism was modified to allow GIFT and the mobile ITS app to send messages to one another, which, in turn, allowed the creation of the gateway needed for GIFT to receive GPS data from a cellular network through a mobile device.

3.2.1. Building a Gateway to a Cellular Network

Since a publicly-facing hosted GIFT instance was set up to allow the mobile ITS app to embed GIFT's tutoring interface, the ITS app had to be set up to both provide a way to view the GIFT webpage and to handle communications with GIFT to perform operations using the mobile device. MIT's Ionic framework was chosen to facilitate this setup process, since it provides a built-in interface that can use embedded web pages to create mobile apps and expose native mobile device features to said web pages (www.ionicframework.com, 2018). By using Ionic, the mobile ITS app's user interface was set up to allow new cadets to pick from several pre-defined courses on the hosted GIFT instance and begin running them at the tap of a button, all within a single mobile app.

With GIFT's tutoring interface now embedded within the mobile ITS app via Ionic, the only remaining hurdle for establishing a gateway to the cellular network was to define the messages that GIFT and the mobile ITS app would send one another and define what operations would be invoked when those messages were received. Since a similar messaging process had already been established for web-based Unity applications, the bulk of the encoding and decoding process used for sending and receiving messages between GIFT and Unity was reused for the mobile ITS app. The existing message types that were used to load and pause Unity applications were re-used to allow GIFT to declare when it wanted to start and pause GPS location tracking, and a message type that was used to display messages in Unity applications was reused to display textual notifications on mobile devices and vibrate them via Ionic. A new "Geolocation" message type was also introduced to allow GIFT to receive updates of the mobile device's GPS location that are requested every second by Ionic. Now that the necessary messaging system and native device calls had been set up between GIFT and the mobile ITS app, the last remaining step to building the land navigation app was to set up assessment logic to allow GIFT to process the GPS location data received from mobile devices.

3.3. Handling Assessment and Pedagogy

Since GIFT's messaging logic had already been modified to handle the GDC location coordinates received from mobile devices' GPS services, the bulk of the changes that needed to be made to GIFT's assessment logic involved simply adjusting condition classes to correctly calculate distances and locations using GDC coordinates. In order to allow the land navigation app to use GPS location data to detect when learners reached certain locations and followed certain paths, GIFT's AvoidLocationCondition, EnterAreaCondition, and CooridorBoundaryCondition classes were all modified to be able to process GDC coordinates just as they would with coordinates from simulated environments. A new PaceCountCondition class was also created to allow GIFT to track learners as they walked a predefined distance and provide feedback once they reported that they had walked that distance.

As for GIFT's pedagogical logic, a need arose for GIFT to be able to display media such as YouTube videos and slide shows as a form of remediation if a learner answered a question improperly while walking through the live scenario. To fulfill this need, a new type of pedagogical strategy was added to GIFT to allow it to display such media during a training scenario, and GIFT's authoring tools were also updated to allow course authors to use this new strategy type. Together, the changes made to GIFT's assessment and pedagogical logic helped build out the course content that was used by the land navigation app as part of its integration into the land nav curriculum.

4. INTEGRATION IN LAND NAV CURRICULUM

With the underlying architectural components in place to support intelligent tutoring in a mobile context, the next task was applying existing module components to an explicit domain use case. With backing from DMI at USMA, the GIFT mobile app was applied to support a Terrain Walk exercise administered to new cadets. In this section, we will give a breakdown of the authoring workflows required to support a Terrain Walk mobile exercise using GIFT, along with examples of those workflows applied for specific training tasks, and a study design applied to gauge the impact of this new type of training intervention and its return on investment.

4.1. Building a Land Navigation Intelligent Mobile App

When designing the learner experience for a mobile Terrain Walk, the following storyboard was generated in collaboration with DMI (see Figure 2). The training event is setup to guide a learner up a path, with activities executed at designated points along the way; as marked by the black dots on the map below. Each activity is designed so that it requires on the spot problem solving through interaction with an individual's available tools (i.e., map, compass, protractor, notebook, etc.) and the natural environment around you. The key objectives instructed include (1) map orientation, (2) terrain association, (3) pace count confirmations, (4) compass checks, and (5) dead-reckoning for building route narratives.

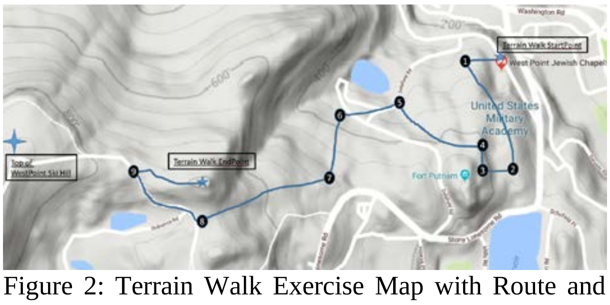

Figure 2: Terrain Walk Exercise Map with Route and Designated Activity Waypoints

When implementing the defined storyboard, there were a number of items to design and instantiate. Primarily, we were concerned with what triggers were required to inform GIFT that an activity should be delivered, and how that activity would be administered so as to collect relevant performance information that can be used to infer competency and guide coaching. The first step is establishing a DKF that outlines the tasks a learner will be asked to perform, the concepts those tasks are designed to target, conditions and standards by which those concepts are assessed, and feedback/remediation material used to coach a learner when errors in performance are recognized. This cognitive task analysis activity establishes a high-level breakdown that designates what needs to be configured within the GIFT authoring tools.

Following this task breakdown, the next activity is configuring the DKF in GIFT's authoring environment. This requires the following: (1) establishing a set of tasks (e.g., plan route from Waypoint A to Waypoint B), (2) configuring start-/end-triggers for each task that activates assessments based on task context, (3) building instructional activity prompts that are displayed when a start-trigger is satisfied, (4) defining the concepts that are assessed within each activity, (5) configuring logic to inform assessments based on the context of the activity, and (6) establishing feedback and remediation when assessments show poor performance and misconceptions. This resulting DKF is used at training runtime to manage the entire learner experience for a given scenario event, with the intent to provide guidance when needed, and to increase challenge when warranted.

One challenge we recognized was the potential for the GIFT mobile app to freeze or crash at a given instance. In this situation, the way in which tasks and DKFs are established is very important. At the first go, we built a single DKF to manage the learner experience. The issue with this approach was each task established in our model was dependent on the other (i.e., one task would initialize on the basis of another task closing). The limitation we recognized was the requirement for a learner to return to the training starting point if the app required a refresh of any sort. To address this, we decided to supplement the entire training experience across six separate courses, where the mobile app would present each course as if they were naturally part of the training intervention (see Figure 3). This created an encapsulated course with six separate interaction lessons. However, each lesson is represented as an individual course, as informed by the GIFT ontology of course terminology.

Figure 3: GIFT Mobile Home Screen for the Land Navigation Terrain Walk Training Experience

With the infrastructure in place to support a sequenced flow of GIFT courses, where each one builds upon the other, the next is establishing the course flow and DKF instantiations for each identified lesson. This requires interaction within two separate GIFT toolsets. The first is establishing the flow of interaction using GIFT's Course Creator (see Figure 4). The Course Creator is used to build a sequence of course objects that dictate the experience. There are a number of course objects available to a GIFT developer, with objects that support the presentation of information, the administration of surveys and tests, the delivery of content, and the management of practice opportunities through scenarios and problem sets.

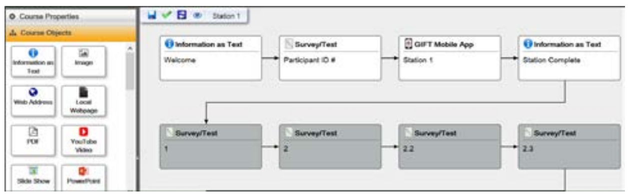

Figure 4: GIFT Course Creator Interface with Station 1 Configuration

With respect to the land navigation use case, each of the six stations depicted in Figure 3 follow the same set of course object sequences to guide the learner experience. This involves providing contextualized information welcoming the learner to the selected station, a survey put in place to collect a learner's participant number, a GIFT Mobile App course object that designates the gateway and DKF to use for training, and contextualized information notifying the learner the station is complete, along with instructions on what to do next (e.g., refresh the app and load the next station).

It's worth noting that there are multiple survey course objects greyed out at the bottom of Figure 4. These surveys are established within the course using GIFT's survey composer (see Figure 5). This tool allows an author to build questions sets with scoring logic that can inform performance states during a training scenario. In terms of course flow, these objects are deactivated with respect to their use within the course as their own separate interactions. This function populates GIFT's survey database, which can be referenced while building a station's oriented DKF, enabling the use of these questions as assessment items triggered by GIFT logic built within the Mobile App Interaction.


Figure 5: GIFT Survey Authoring Tool Used to Build Terrain Walk Assessment Items Used During Training

4.1.1. Configuring Interaction within GIFT's Domain Knowledge File (DKF)

With a set of course objects in place, and surveys built to support assessment practices, next is configuring the mobile app. The primary interactive component of each lesson is encapsulated within the GIFT Mobile App course object. To build the mobile app object, two things are required: (1) establishing the training environment/gateway that will dictate the data sources captured during the training event (i.e., GPS location in this case), and (2) establishing the DKF called upon when the course object is activated. For this specific use case, six total DKFs were required, as the course was dissected into six independent interactions. To build each DKF, GIFT's Real-Time Assessment authoring environment was used.

The first component to building a DKF to support land navigation training in a mobile context is establishing scenario properties that are referenced by tasks (i.e., potential start triggers) and assessment conditions for task-related concepts. In this instance, we populate designated GPS coordinates provided in the storyboard that are used by GIFT to drive the learner experience (see Figure 6).

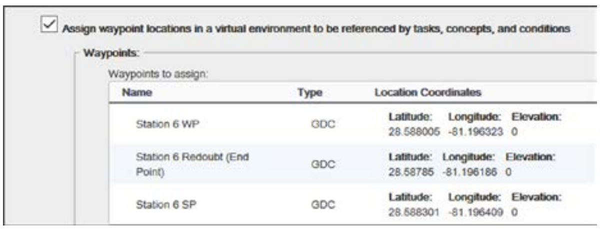

Figure 6: Waypoint References Populated in the DKF's Scenario Properties for Station 6 Start Point (SP), Station 6 Waypoint (WP), and Station 6 End Point (EP)

With the scenario properties populated, the next requirement is establishing the tasks a learner will execute, the concepts being assessed within each task, and the conditions used for assessing performance across the populated concepts. One mechanism that needed to be addressed was building in location oriented triggers that are used to notify the learner of an activity, along with providing the logic to present questions. As seen across Figures 7 and 8, there are two tasks represented within the DKF schema: (1) Station 6 SP Location and (2) Station 6 SP. The first task (see Figure 7) is associated with tracking a learner's movement and notifying GIFT when that learner reaches a designated location. The 'Avoid Area' condition class was used to support this function. Based on the configuration, if the learner comes within 5 meters of the defined Station 6 SP (see Figure 6), GIFT creates a state transition message that signifies the condition has been met. This state transition property is then linked to an instructional strategy selection, where GIFT executes a string of actions when the established condition is recognized. In this instance, when the learner enters that waypoint, GIFT presents station instructions, gives the learner 10 seconds to ready the content, and then presents the first question that requires the learner to execute an activity.

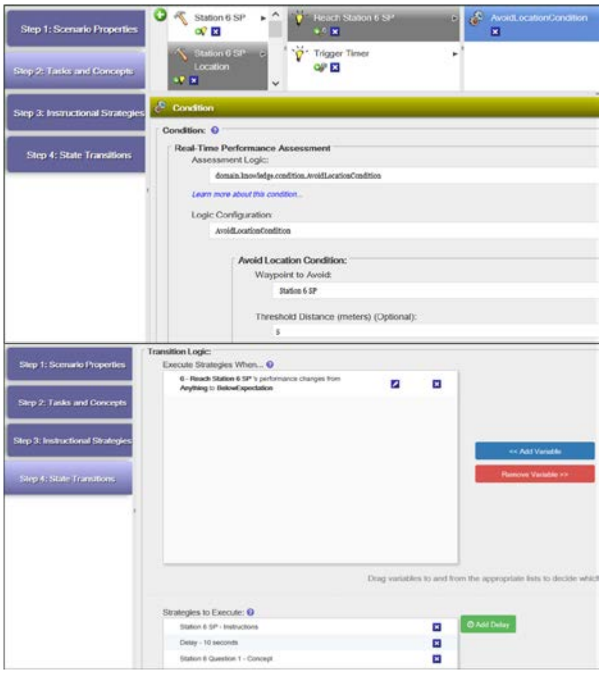

Figure 7: The task and concept structure in the DKF for tracking location, creating state transition messages when concept conditions are met, and configuring what instructional strategies to execute when the transition is observed

The next task represented in the DKF associates with question-based assessments that inform performance on a set of concepts (see Figure 8). Within this representation includes: (1) the designation of a pre-authored survey to present that was established to assess a concept and infer performance, (2) state transitions that dictate pedagogical approaches when certain performance states are observed, and (3) a set of feedback strings or media content to display when the defined state transition is observed.

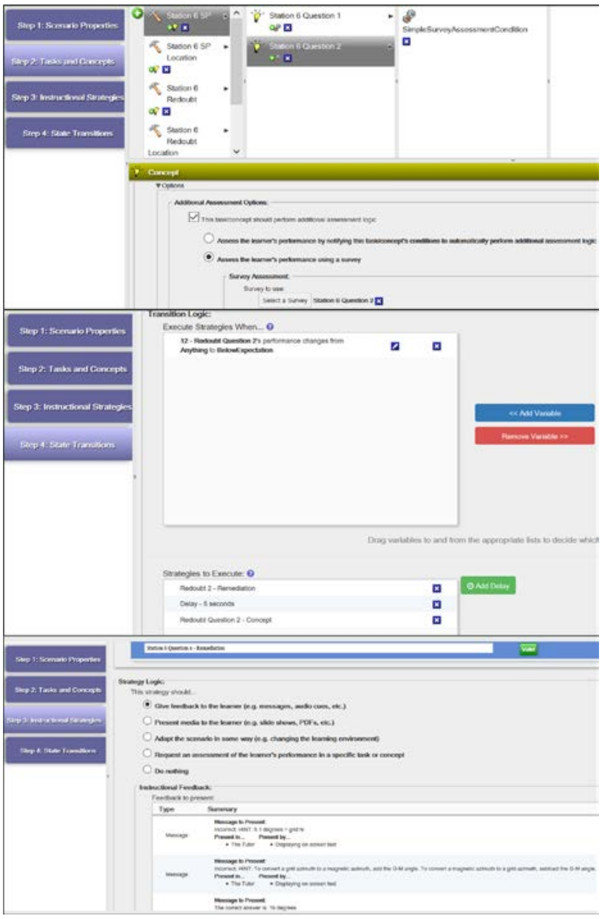

Figure 8: The task and concept structure in the DKF for presenting a question, state transition messages based on question performance, and what instructional strategies to execute when the transition is observed

In this instance, because the learner reached the SP, GIFT presents instructions and then displays a question. The question is designed to require a learner to use the tools provided to execute a task contextualized by their surrounding environment. For Station 6, a waypoint is provided to the learner and they are required to plot the point on a map, measure the distance to that point from their current location, and calculate an azimuth they will hold when executing the defined route. This is accomplished by using a map, protractor, and compass to gather the necessary information. Each question in GIFT is designed to gather input on the required concepts needed to effectively perform the task, with specified feedback that targets misconceptions and erroneous inputs.

The question referenced in Figure 8 requires the trainee to input the azimuth from their location to the assigned waypoint. If the answer provided by the learner falls outside the acceptable threshold, a below-expectation transition state is produced, which in turn triggers a feedback strategy. In this instance, the feedback strategy is designed to present a single string of text, as defined by the instructor, to the learner, followed by a re-presentation of the question. The feedback strategy allows for escalation of detail, where an author can build in layers of feedback if the question is continually answered incorrectly. While a single example has been reviewed in the previous paragraphs, this process is required for across all stations, all tasks, and all associated concepts. As an example of the learner interface on the app, Figure 9 shows a screenshot of GIFT Mobile presenting a question regarding the distance of a landmark from the top of a lookout. If the answer is inferred as incorrect, feedback logic is presented, followed by a new attempt to answer the question.

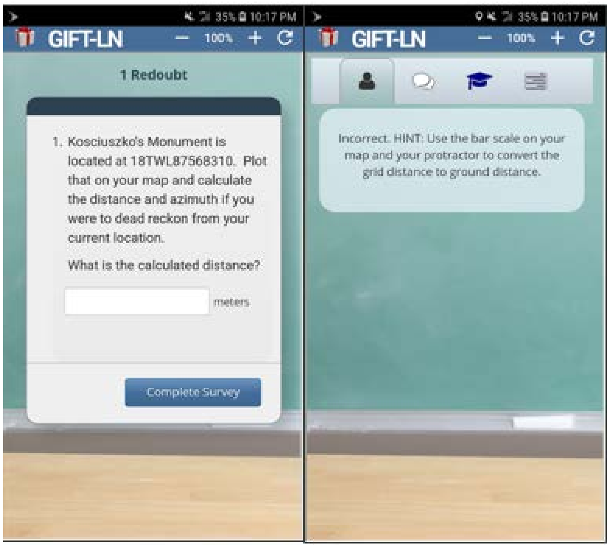

Figure 9: GIFT's mobile interface a learner interacts with while executing the land navigation Terrain Walk exercise

The process described in this section was repeated across all tasks and concepts for all associated stations outlined in the Terrain Walk storyboard. With a completed app in place, DMI supported the execution of a pilot study to gauge the reaction and impact of this form of technology insertion into the traditional training cycle.

4.2. Training Pilot Study

With a fleshed out prototype of the GIFT Mobile App, DMI approved an initial pilot study to gauge the impact of this new technology when compared to traditional approaches. This provides a control condition we can compare the mobile app interaction and performance outcomes against. While the control does not provide explicit measures we can use to compare performance outcomes, we administered pre- and post-tests to all new cadets that assessed their ability to execute the foundational skills the Terrain Walk targets. We also have access to the new cadet performance outcomes on the day five qualification event.

In the experimental treatment, a group of twenty new cadets were randomly selected from their company (i.e., approximately 160 new cadets) and outfitted with a GIFT configured Android device. The participants were provided a 2.5 hour window to complete as much of the course as possible, with the route extending over 2.5 miles. Over the course of a four-day window, we consented 142 new cadets. We are currently working with the data to extract and prep it for analysis and model builds. The outcomes will be highlighted in future reports.

5. CONCLUSION

In this paper, we review the development and authoring processes required to make a mobile GIFT app focused on land navigation training. To build the resulting prototype, modifications to the GIFT architecture were required to support interaction across cellular networks, while interfacing with smartphone displays. The design requirements and authoring workflows were described for a specific use case, but each process described is generalizable and can extend to multiple domains that incorporate movement and interaction in a live operational environment. With a functioning mobile GIFT app, a pilot study was designed to gauge the technology impact. Future efforts will focus on app enhancements and modifications based on the weaknesses identified in the pilot-test event. In addition, the app will be expanded to support pedagogy during the practice opportunities provided on a specific land navigation course where a learner is left to execute on their devices.

ACKNOWLEDGMENTS

We would like to thank the United States Military Academy (USMA) and the Department of Military Instruction for supporting these efforts. Specifically, Captain Luke Plante was critical in executing the data collection event.

REFERENCES

Amory, A., Naicker, K., Vincent, J., Adams, C., 1999. The use of computer games as an educational tool: Identification of appropriate game types and game elements. British Journal of Educational Technology, 30(4), 311-321.

Bratt, E. O., 2009. Intelligent tutoring for ill-defined domains in military simulation-based training. International Journal of Artificial Intelligence in Education, 19, 337-356.

Bruner, J. S., 1966. Some Elements of Discovery. In: L. S. Shulman and E. R. Keislar, eds. Learning by discovery: A critical appraisal. Chicago, IL: Rand McNally, 101-113.

Cuevas, H. M., Fiore, S. M., Bowers, C. A., Salas, E., 2004. Fostering constructive cognitive and metacognitive activity in computer-based complex task training environments. Computers in Human Behavior, 2006, 225-241.

Ericsson, K. A., Ward, P., 2007. Capturing the naturally occurring superior performance of experts in the laboratory: Toward a science of expert and exceptional performance. Current Directions in Psychological Science, 16(6), 346-350.

Greitzer, F. L., Kuchar, O. A., Huston, K., 2007. Cognitive science implications for enhancing training effectiveness in a serious gaming context. ACM Journal of Educational Resources in Computing, 7(3).

Hermann, G., 1969. Learning by discovery: a critical review of studies. Journal of Experimental Education, 38(1), 58-72.

Johnson, D. W., Johnson, R. T. and Smith, K. A., 1991. Active Learning: Cooperation in the College Classroom. Edina, MN: Interaction Book Company.

Menaker, E., Coleman, S., Collins, J., Murawski, M., 2006. Harnessing Experiential Learning Theory to Achieve Warfighting Excellence. Proceedings of the Interservice/Industry Training, Simulation & Education Conference, Orlando (Florida, USA).

Nicholson, D., Fidopiastis, C., Davis, L., Schmorrow, D., Stanney, K., 2007. An Adaptive Instructional Architecture for Training and Education. Proceedings of the Human Computer Interaction-International (HCII) 2007, Berlin (Germany).

Oser, R. L., Cannon-Bowers, J. A., Salas, E., Dwyer, D. J., 1999. Enhancing performance in technology-rich environments: Guidelines for scenario-based training. In: E. Salas, ed. Human/technology interaction in complex systems. Stamford, CT: JAI Press, 175-202.

Pine, A., 2009. Simulation: The (almost) real thing. Proceedings Magazine, 135(12/1), 282.

Salas, E., Priest, H. A., Wilson, K. A., Burke, C. S., 2006. Scenario-based training: Improving military mission performance and adaptability. In: A. B. Adler, C. A. Castro & T. W. Britt, eds. The psychology of serving in peace and combat operational stress (Vol. 2). Westport, CT: Greenwood Publishing Group, Inc.

Sottilare, R., Goldberg, B., Brawner, K. W., Holden, H., 2012. Modular Framework to Support the Authoring and Assessment of Adaptive Computer-Based Tutoring Systems. Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando (Florida, USA).

VanLehn, K., 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197-221.

Waldman, A., 2009. Serious gaming. Available from: http://www.ngaus.org/content.asp?bid=14851 [accessed 15 July 2018]

Wisher, R. A., Macpherson, D. H., Abramson, L. J., Thorton, D. M., Dees, J. J., 2001. The Virtual Sand Table: Intelligent Tutoring for Field Artillery Training: U.S. Army Research Institute for the Behavioral and Social Sciences.

BENJAMIN GOLDBERG is a member of the Learning in Intelligent Tutoring Environments (LITE) Lab at the Natick Soldier Research and Development Center's Simulation and Training Technology Center in Orlando, FL. He has been conducting research in intelligent tutoring for the past ten years with a focus on adaptive learning in simulation-based environments and how to leverage Artificial Intelligence tools and methods to create personalized learning experiences. Dr. Goldberg holds a Ph.D. from the University of Central Florida in Modeling & Simulation.

NICHOLAS ROBERTS is a software engineer at Dignitas Technologies on the Generalized Intelligent Framework for Tutoring (GIFT) project with around five years of software development experience. His work primarily consists of implementing many of GIFT's user interfaces as well as supporting and adding features to GIFT's web applications, including the dashboard, the course creator (i.e. the Gift Authoring Tool), and course runtime interface (i.e. the Tutor). Nicholas also often provides troubleshooting support for the GIFT portal's forums (www.GIFTTutoring.org), particularly for issues surrounding GIFT's user interfaces, integrating new technologies with GIFT's web applications, and all things related to the Google Web Toolkit (GWT). He acted as the lead GIFT developer for implementing the communication protocol between the Android Phone and GIFT and added the logic needed to support performing assessments based on the phone's geographical position.

W. GABE POWELL is a U.S. Army Captain (Ret) with prior faculty service under the United States Military Academy's Geospatial Information Science Program. He also served as faculty under the Department of Military Instruction and was responsible for planning and execution of land navigation training for incoming new cadets. Captain Powell's military career is focused in applied and tactical intelligence. Since enlisting in the U.S. Army in 2004, Powell has completed missions in humanitarian relief, counterterrorism, and nuclear response. He provided Subject Matter Expertise throughout the execution of the land navigation mobile app research efforts.

ELYSE BURMESTER is a research assistant for Dignitas Technologies currently supporting the Adaptive Training team within the U.S. Army Natick Soldier Research and Development Center (NSRDEC) in Orlando, Florida. She is in her first year of experience within U.S. Army research and development agencies. She holds a Bachelor's degree in Political Science with a minor in Psychology from Florida Gulf Coast University. She has interest in adaptive training, ITS architectures, data structures and machine learning.

A HYBRID MACHINE LEARNING APPROACH TO AUTOMATED SCENARIO GENERATION (ASG) TO SUPPORT ADAPTIVE INSTRUCTION IN VIRTUAL SIMULATIONS AND GAMES

Robert A. Sottilare
US Army Natick Soldier Center
robert.a.sottilare.civ@mail.mil

ABSTRACT
This paper examines machine learning methods to automatically generate a large number of child scenarios from a small number of parent scenarios in support of adaptive instruction conducted in virtual simulations and game-based platforms. Adaptive instructional systems (AISs) include Intelligent Tutoring Systems (ITSs), intelligent mentors, recommender systems, personal assistants, and intelligent instructional media. AISs attempt to tailor instruction for individuals and teams based on their learning needs (e.g., knowledge or skill deficiencies), goals, and preferences. This often requires much more content than current non-adaptive systems which provide one or a very limited set of training scenarios to address a given set of learning objectives. The goal of the research described in this paper is to reduce the authoring burden for developing a large number of unique and relevant training scenarios. The methodology presented also ranks the resulting scenarios with respect to a set of author-specified learning objectives and learner/team competency in the domain of instruction. The unique contributions of this paper are tied to its hybrid machine learning approach, and consideration for both learning objectives and learner/team competency in automatically ranking generated scenarios.

Keywords: adaptive instruction, adaptive instructional systems (AISs), automated scenario generation (ASG), combinatorial optimization search (COS), evolutionary scenario generation (ESG), genetic algorithm (GA), Intelligent Tutoring System (ITS), machine learning, novelty search, ranking algorithm, reinforcement learning

1. INTRODUCTION

Adaptive instructional systems (AISs) are a class of intelligent, machine-based tools that "guide learning experiences by tailoring instruction and recommendations based on the goals, needs, and preferences of each learner [or team] in the context of domain learning objectives" (Sottilare & Brawner, 2018a; Sottilare & Brawner, 2018b; Brawner & Sottilare, 2018). AISs may include technologies like intelligent tutoring systems (ITSs), intelligent mentors, recommender systems, personal assistants for learning, and intelligent instructional media.

In late 2017, the Institute for Electrical and Electronics Engineers (IEEE) Learning Technologies Standards Committee (LTSC) under the auspices of the IEEE Standards Association established an AIS study group to examine opportunities for potential standards to lower the entry and maintenance costs associated with AISs and the AIS study group identified authoring (development) as a major barrier to the adoption of AISs. The AIS authoring process can be divided into two sub-processes: 1) developing or finding appropriate content (often called curation), and 2) sequencing or aligning content with learning objectives (sometimes called building or configuring depending on the authoring tool).

Specifically, the AIS standards study group identified the skill and cost of authoring these systems as very high, usually requiring highly technical individuals with expert programming skills to develop and maintain these systems. Compounding this problem is the fact that AISs often require significantly more content (as much as 2 or 3 times) than non-adaptive systems since each adaptation (tailored instructional sequence) requires new content, and each remediation also requires new content.

We are suggesting that the authoring barrier might be reduced by automating as much of the authoring process as possible. Ideally, we would want to fully automate the entire AIS authoring process, but are taking the approach to solve one problem at a time beginning with the complex problem of automated scenario generation (ASG; Zook, et al, 2012) which can greatly expand the content choices for adaptive instruction offered by authoring tools like the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg, & Holden, 2012; Sottilare, Brawner, Sinatra, & Johnston, 2017), the Cognitive Tutor Authoring Tool (CTAT; Aleven, McLaren, Sewall, & Koedinger, 2006), the Authoring Software Platform for Intelligent Resources in Education (ASPIRE; Mitrovic et al, 2006) and other AIS authoring platforms.

The goal of ASG is to create training scenarios for domains that vary in their complexity, definition, and dynamics (Sinatra & Sottilare, 2016), and are ranked by their relationship with specified learning objectives. The basic idea is to automatically create significantly different scenarios where all the variables in the scenario are allowed to vary maximally resulting a large number of training situations available to support tailored instruction. Of course, not all the scenarios created would be relevant, doctrinally correct or even possible in the real-world. A mechanism is needed to rank their relevance or fitness with respect to a set of learning objectives and the competency of the learner in performing the assigned task.

By way of example, we have selected a *room clearing task* under varying conditions to illustrate the functional aspects of ASG and how it might work for military training. Usually, an instructional developer would be responsible to handcraft each scenario using a scenario editor specific to the game/simulation being used and their expertise in the domain of instruction. A military or law enforcement squad or fire team would usually train to master the task of entering and clearing a room of any hostiles. ASG is critical to providing both challenging and doctrinally correct scenarios for adaptive team training.

The next section of this paper explores the scope of the ASG problem space by way of defining terms and describing the process associated with a genetic algorithm (GA) approach.

2. SCOPING THE PROBLEM OF ASG

As part of examining the ASG problem space, we thought it would be useful to provide a few definitions to help shape the scope of our discussion:

- ***Scenario*** - a process in which a learner or learners interact within an environment over a sequence of events which introduce and/or exercise a set of skills defined by a set of learning objectives
- ***Fitness Function*** – criteria used to assess how close a scenario is to achieving a set of defined objectives
- ***Scenario Generator*** - a computational system that solves the problem of producing a set of viable scenarios given knowledge about their attributes and their alignment with the fitness criteria
- ***Initial Population*** – an initial set of scenarios that adequately represent a set of targeted learning objectives and are used to generate future scenarios through some machine learning technique
- ***New Population*** – a resulting set of scenarios automatically generated that are more closely aligned with the fitness criteria than their parents; a scenario's fitness is determined by the weighted linear sum of all evaluation functions

The ASG problem space can be distilled into three distinct challenges: 1) how to insure sufficient variation in the parent population so these traits are passed to subsequent generations; 2) how to promote sufficient variation of complexity and tailoring in the subsequent generations/new populations; and 3) defining the fitness criteria to evolve and rank a population of new scenarios that support specified learning objectives, support goals, preferences, and learning needs of individuals or teams, and are realistic.

Given the definitions and challenges described, we can now concentrate on describing a generalized process for ASG using GAs (Figure 1). We chose to use GAs based on their flexibility in addressing a variety of tasks, their ability to cover the search space, and their ability to address the three challenges we identified. In the next section of this paper, we explore three approaches to developing an ASG capability using GAs.

3. EXAMINING POTENTIAL GA APPROACHES

According to Zook et al (2012), a genetic algorithm usually starts with a population of randomly generated potential scenarios and attempts to modify and/or combine aspects of different scenarios within the population to improve the fitness of the next generation of scenarios according to a given fitness function. A GA is "a search heuristic that is inspired by Charles Darwin's theory of natural evolution" (Mallawaarachchi, 2017) that generates a pool of candidate solutions called a population.

GAs have an advantage over gradient based methods which may trend toward local optima for many complex real-world domains. GAs have the ability to provide a large number of usable (good enough) solutions relatively quickly (Wikipedia, 2018). GAs are also relatively easy to implement and resolve to a solution in most cases. For ASG, this makes them a more attractive choice over other approaches (e.g., deep reinforcement learning, artificial neural networks) which may be difficult to implement.

For our exploration of ASG using GAs, we will use a "*room clearing training task*" as a basis for the examination of three machine learning approaches that exploit genetic algorithms:

- Brute Force Search
- Novelty Search
- Combinatorial Optimization Search

3.1. Brute Force Search Approach

A brute force search (also known as an exhaustive search) solves the generation problem by systematically enumerating *all possible candidates* for the solution and checking whether each candidate satisfies the problem's statement (Wikipedia, Brute-Force Search, 2018). Depending upon the number and type of variables, how we decide to implement the GA for our room clearing task could be complex or very simple. Our example task discussed in Section 4 of this paper is very simple in order to illustrate the principles and process of implementing a GA.

Brute force searches are the easiest to implement, and always find a solution if one exists. However, as the number of candidate solutions grows, the search time also grows rapidly. "Therefore, brute-force search is typically used when the problem size is limited, or when there are problem-specific heuristics that can be used to reduce the set of candidate solutions to a manageable size. The method is also used when the simplicity of implementation is more important than speed" (Wikipedia, Brute-Force Search, 2018).

3.2. Novelty Search

Fitness functions for genetic algorithms are typically goal-focused. The goal in ASG is to understand the alignment of candidate scenarios with specified learning objectives. An exception to this is novelty search. *Novelty search* uses a fitness function to promote behavioral novelty instead of attempting to conduct a search through the use of a static objective or set of objectives (Lehman, 2012). Since the goal of this search is to identify unique candidates, the result is more likely to include candidates outside of what might be normally acquired through a static objective search or in systems where the number of candidates is limited. Conversely for ASG, large numbers of initial candidates (parent scenarios) are likely to yield a more diverse set of child scenarios resulting in some non-viable scenarios.

Genetic algorithms attempt to satisfy a criteria set by the fitness function so they generally do not identify optimal candidates and with the exception of novelty search do not tend to cover the entire search space. Novelty search in its attempt to identify all unique candidates does tend to cover the more remote areas of the search space often left uncovered by other GA approaches.

3.3. Combinatorial Optimization Approaches

Combinatorial Optimization uses a scenario generation approach to deliver the requisite diversity and quality of scenarios while tailoring the scenarios to a particular learner's needs and abilities. This type of optimization includes an eight step process illustrated in Figure 1 and discussed in detail in Sections 3.1.-3.6 (Shiffman, 2012).

- Step 1: define fitness criteria
- Step 2: create an initial population of N scenarios
- Step 3: assess the fitness of each individual within the population based on the fitness criteria until stop criteria is met, then go to step 8
- Step 4: create a mating pool based on fitness scores and select pairs for reproduction
- Steps 5 & 6: reproduce N times through cross-over and mutation and add each child to the new population
- Step 7: replace the old population with the new one and return to step 3
- Step 8: print results and terminate

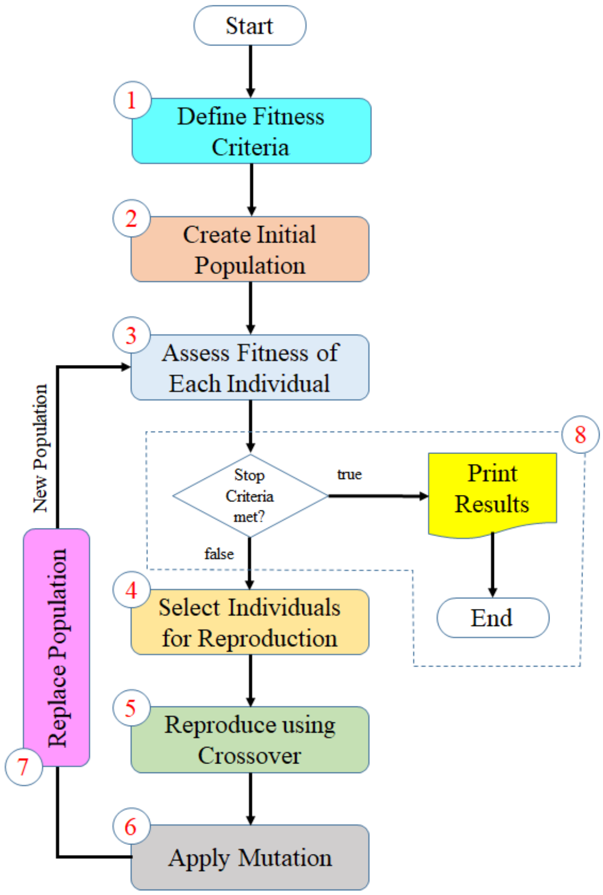

Figure 1. Combinatorial GA Approach

In the next section of this paper, we discuss how we might apply and vary the GA approaches reviewed above to support ASG for our example task, clear rooms training task.

4. COMBINATORIAL OPTIMIZATION WITH NOVELTY SEARCH

In this section, we layout a process for combinatorial optimization with novelty search. The steps discussed below are applicable to a variety of task domains and we provide an example task to illustrate its use. The example task, an understanding of its elements and how these elements relate to a specified fitness function are critical to reusing this process for other domains.

For our example task we have chosen a "*clear rooms training task*" (US Army, 2007) which is usually performed by squads of dismounted soldiers on patrol. Since the room clearing task is a psychomotor task, we refer to the GIFT authoring tools which use a psychomotor task model based on theories advanced by Dave (1970), Simpson (1972), Harrow (1972), and Romiszowski (1999) as adapted by Brown, Bell & Goldberg (2017). The goal is to have the team practice and demonstrate a level of proficiency where automaticity and fluid motion are the norm. In a game-based training environment, the focus is more on the cognitive aspects of the task and mastering the interface, but in a fully immersive virtual environment, the focus is on mastering the physical aspects of the task, interaction with the environment, and interaction (e.g., communication and coordination) with other members of the team.

All military training scenarios describe the task, the conditions under which each task is conducted, and the standards or measures of successful performance. The standards or measures of success for our room clearing task include:

- Enter the room quickly and smoothly
- Clear the doorway immediately
- Remain within arm's reach of another squad member
- Secure room by neutralizing any enemy present
- Maintain sufficient force to defeat any enemy counterattack and continue operations

Note that even an identical environment (e.g., same room layout and same threats at the same locations) could result in a different scenario based on the squad's decisions and performance. The simple decision of entering the room at a different location can impact the sequence of events to follow. If we examine the dynamic elements (e.g., skill events, environment, and constraints) of our room clearing training example, we find that our scenarios can vary by type, sequence, length, and outcome of events, but can also vary in complexity by changes to the environment (e.g., threats or the physical configuration of the building) and the number and type of constraints (e.g., rules of engagement or presence of non-combatants).

Next, we describe eight essential steps in the GA process described below and shown in the context of Figure 1. We have modified these to fit our example training task and to overcome our defined set of challenges:

- define our fitness criteria
- insure sufficient variation in the parent population
- promote sufficient variation of complexity and tailoring in the child population

For our technical approach to ASG (described in the eight steps below), we have chosen to primarily use a combinatorial optimization approach since it would inefficient to pursue a brute force approach for more complex scenarios than our example task. This will allow us to represent more complex domains in the future with the same GA process. We have also chosen to substitute a Novelty search in Step 2 (create initial population) to provide a more full representation of the search space and provide a higher degree of variability in subsequent generations.

The resulting combinatorial optimization with novelty search process for ASG is based on a Darwinian model of evolution through natural selection and genetic variation:

- Step 1: define fitness criteria
- Step 2: create an initial population of N scenarios has two substeps:
  - 2a. select 3-4 scenarios that vary across the variables selected for the fitness criteria
  - 2b. use Novelty search to expand this population to N unique scenarios using single point crossover and single point mutation
- Step 3: assess the fitness of each individual within the population based on the fitness criteria until stop criteria is met, then go to step 8
- Step 4: create a mating pool based on fitness scores and select pairs for reproduction
- Steps 5 & 6: reproduce N times through cross-over and mutation and add each child to the new population
- Step 7: replace the old population with the new one and return to step 3
- Step 8: print results, output scenario editor file and terminate

4.1. Step 1: Define Fitness Criteria

The first and most important challenge is to define our fitness criteria such that scenarios that more closely align to our learning objectives are ranked higher than those that are more loosely aligned with the learning objectives.

In scoping the ASG problem space, it is necessary to understand the relationship between task learning objectives and attributes of potential solutions in the initial population. For our task, a squad will be trained to clear one or more rooms in a building in a virtual simulation (e.g., Virtual Battle Space). The room clearing task involves many coordinated behaviors, but the learning objectives or standards for the squad can be distilled into five essential assessments defined previously in this section.

The complexity of the task can vary and scenarios that account for varying complexity can be tailored to the competency (experience or prior knowledge) of the team members. Each of these task learning objectives and thereby any associated scenario may be complicated by the:

- Size and shape of the room
- Number of armed enemy forces present
- Number of non-combatants present
- Obstacles at the doorway or in the room

Given this task and varying complexity, we represented any given solution in the population of possible scenarios as a four digit integer where each integer varies from 0 to 9:

- Size and shape of the room - where 0 = simplest room (e.g., a small rectangular room) and 9 =

most complex room (e.g., large room with interior corners and multiple doorways)

- Number of armed enemy forces present – where 0 = no enemy forces present, and 9 = 9 enemy forces present
- Number of non-combatants present – where 0 = no non-combatants present, and 9 = 9 combatants present
- Obstacles at the doorway or in the room – where 0 = no obstacles and 9 = 9 obstacles present

Assuming it is physically possible to fit 9 enemy forces, 9 non-combatants and 9 obstacles in the smallest room, this would mean we can generate up to 10^4 scenarios using a genetic algorithm approach as shown in Figure 2.

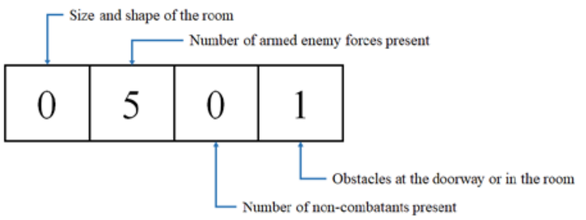

Figure 2. Population Representation Schema

For simplicity, we have elected to measure the complexity of a scenario by summing the four genes or attributes that make up a scenario. For example, the complexity for the chromosome or potential solution shown in Figure 2 would be 6 (0 + 5 + 0 + 1) where 0 would is the lowest complexity and 36 the highest. If we align the complexity of scenario with the domain competency of the team, we would have alignment within Vygotsky's (1987) Zone of Proximal Development (ZPD; Figure 3). This alignment will help maintain engagement and positive affect during the training process.

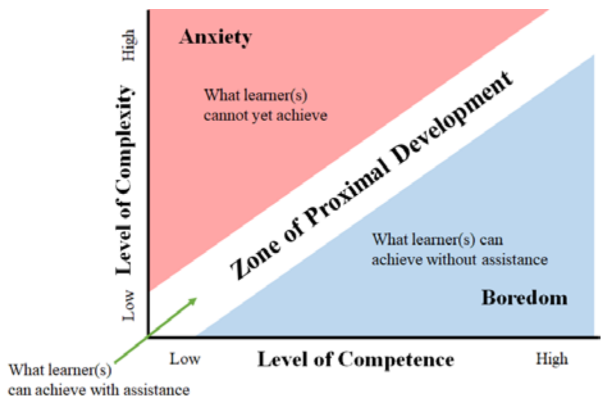

Figure 3. Zone of Proximal Development

We could then use this as a fitness criteria by comparing the complexity of the scenario and domain competency of the team. Defining the team domain competency in four intervals between 0 and 36 provides the following distribution for team competency:

- Expert (28-36)
- High Skills (19-27)
- Moderate Skills (10-18)
- Low Skills (0-9)

We assume that some long term model of the team's domain competence, a pretest, or subject matter assessment would determine where any particular team would fall on this competency scale, but the fitness function for this example task domain would be:

Fitness = domain competency – scenario complexity

For example, a moderately skilled team with a domain competency score of 17 would be sufficiently challenged by a scenario with a complexity of $17 \pm \sigma$. Assuming $\sigma = 4$, then any scenario in the range of 13-21 would be at an appropriate level of complexity for that particular team. While this example may be overly simplified, it does illustrate the process which could be applied in more complex training domains. If we wished to generate the 10 most appropriate scenarios for this team, we would rank them from lowest difference to highest difference.

4.2. Step 2: Create an Initial Population

The next step in the process is to generate a set of individual scenarios (solutions or chromosomes) which comprise the population. Parameters are represented in the chromosomes as variables are known as genes. Normally, the initial population is generated randomly, but it is critical that sufficient variability is represented in this initial population or the GA search will produce limited results. We have chosen to start with a limited set of four scenarios and then stretch the variability of the population through Novelty search. In this way, we might find sufficient variety in future generations if variability is also represented in the initial population.
For example, an initial population should contain at least one scenario for each of the levels of complexity (easy, moderate, and hard) could be expanded using Novelty search. In our case, we aligned domain complexity intervals with the competency intervals defined in Step 1. Randomly using 0150 (low complexity), 5273 (moderate complexity), 9911 (high complexity), and 4997 (very high complexity) as a seed population for Novelty search will result in several solutions that represent a large portion of the scenario complexity required for future generations. It also allows us to expand our approach to represent other learner/team attributes beyond complexity and competency. The resulting unique set of scenarios will be sufficient to act as an initial population for a combinatorial optimization approach (e.g., crossover and mutation) in subsequent steps discussed below.

4.3. Step 3: Assess the Fitness of Individuals

The fitness function determines the suitability of an individual scenario as a potential solution. The candidate solutions in the population are assessed with respect to the learning objectives which we used to determine the variables in the GA search and matched to scenarios aligning with the competency level of the team.

4.4. Step 4: Select Individuals for Reproduction

In our problem space, ASG, the GA selects the fittest individual scenarios in the current generation to produce offspring for the next generation of the population using a fitness function. In the selection step the goal is to pair the fittest individuals to let them reproduce and pass their genes to the next generation. Some number of pairs (two or more) are selected where the probability of selection of an individual scenario for reproduction is based on its fitness score. Again, assuming a team competency of 17, we selected the top four scenarios in terms of fitness in our first generation resulting from Novelty search for our example task:

- 1772
- 1952
- 1970
- 0773

4.5. Step 5: Reproduce using Crossover

Genetic algorithms are usually used to find solutions meeting the fitness criteria by employing operators like crossover. A single crossover point in the chromosome is chosen and the genes prior to the crossover point are exchanged between the pair of scenarios. For our example task, using the four fittest scenarios defined in Step 4, we might see a pairing between 1772 and 1952 resulting in 1972 and 1752 (Figure 4).

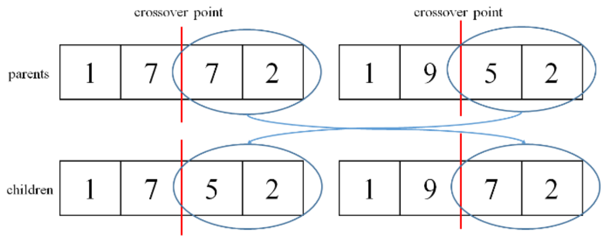

Figure 4. Reproduction Using Crossover (also known as Recombination)

4.6. Step 6: Apply Mutation

In a percentage of the new individuals formed by the crossover reproduction, a random gene is selected for change (single point mutation). Mutation randomly alters a parameter of a randomly chosen event in the scenario and then reevaluates the mutation to determine if its fitness has improved. Mutation is critical step in maintaining diversity in the new population. For our example task, we might see a random mutation that changes 1972 to 8972. While this might seem inefficient since the fitness of the scenario went from 0 to 9, overall mutation infused the new population with greater diversity while the average fitness of the population will continue to optimize.

Note that in addition to crossover and mutation, other less used operators include addition and deletion (Mitchell, 1996; Zook et al, 2012). Addition inserts a random event into the scenario at a random location and then reevaluates the resulting candidate scenario to determine if its fitness has improved. Deletion removes a random event from the scenario. For simplicity, we elected to apply only crossover and mutation operators for our example task.

4.7. Step 7: Replace Old Population with New Population

In this step, we discard the old population in favor of the new population created using crossover and mutation. It is important for the author to select a large enough number of iterations to allow the average fitness score of each new population to trend toward some optimal value or plateau.

4.8. Step 8: Terminate

In this step, we determine when to terminate the ASG process and output the results. The ASG process may continue until a termination trigger is reached:

- a candidate solution is identified that satisfies some minimum criteria
- a fixed number of generations is reached
- an allocated amount of time has elapsed
- candidate solutions reach a level of fitness where they plateau (no significant change)

The output of the ASG process is two-fold:

- Printed list of scenarios (e.g., 1772) with their associated fitness scores
- Scenario editor input (digital file compatible with common scenario editors for games and immersive virtual environments used for training

5. RESULTS

The pseudo code below represents the resulting combinatorial optimization with novelty search GA used to generate scenarios based on the example training task of room clearing:

1. START: Set competency target = 17 with *fitness = domain competency – scenario complexity* and *fitness goal = 0*
2. Selection: randomly 3 initial scenarios with significantly different complexity scores (results = 0342, 5172, 9145)
3. Novelty Search: expand initial population to 10 unique scenarios using single point crossover and single point mutation (10%) to create 7 additional new scenarios (results = 0342, 5172, 9145, 0372, 5142, 0345, 9142, 5145, 9172, 7372)

4. Compute fitness of each individual scenario (results for generation 0)
    a. 0342 fitness = abs(17-9) = 8
    b. 5172 fitness = abs(17-15) = 2
    c. 9145 fitness = abs(17-19) = 2
    d. 0372 fitness = abs(17-12) = 5
    e. 5142 fitness = abs(17-12) = 5
    f. 0345 fitness = abs(17-12) = 5
    g. 9142 fitness = abs(17-16) = 1
    h. 5145 fitness = abs(17-15) = 2
    i. 9172 fitness = abs(17-19) = 2
    j. 7372 fitness = abs(17-19) = 2
5. REPEAT
    a. Selection for mating pool – based on fitness and stochastic universal sampling (Baker, 1987)
    b. Crossover
    c. Mutation (10%)
    d. Compute fitness
6. UNTIL population has converged
7. END

6. CONCLUSIONS, CHALLENGES AND NEXT STEPS

We presented a process and schema for applying a hybrid (Combinatorial Optimization with Novelty Search) GA approach to the automated authoring of scenarios for games and immersive virtual environments. The process provided wide variability for the resulting scenarios that were aligned to author specified learning objectives and learner/team competency to support adaptive instruction.
The ASG approach in this paper is applicable to a broad number of domains in digital training environments. While we focused this application of GAs to ASG for games and virtual simulations, we also see application of this process to live simulations (e.g., mission rehearsal).
The GA approach to ASG described herein benefits greatly from more specific domain knowledge resulting in better objective values. Of course it takes some time for a person to define schema and to incorporate this specific knowledge in each new domain, so this process is not fully automatic, but can be for the end user once the schema and fitness criteria are defined. A likely next step is to create an author dashboard for unit commanders and subject matter experts to lead them through the process of developing learning objectives and critical variables as input to the ASG process defined herein.
The primary challenges defined were three-fold: 1) generating a large number of feasible scenarios which cover a large portion of the search space for variables like complexity (e.g., easy, moderate, and hard scenarios) and which cover the highest percentage of the learning objectives, 2) ranking scenarios in order of relevance to a set of author-defined learning objectives, and 3) providing output from the ASG process which is compatible with scenario editors for both games (e.g. Virtual Battle Space) and immersive virtual environments used for military training. The first two challenges have been met by the process described in this paper along with alignment between learner/team competency and scenario complexity ala Vygotsky's ZPD. This tailoring will enhance the relevance of the scenario for the learner/team and thereby enhance engagement and learning.
The third challenge has not been met, but is simply a mechanical translation of the scenario code to a format that can be understood by scenario editors for games and virtual simulations. The delay in meeting this challenge is based on understanding what games/simulations will have the highest use and thereby the most need for the ASG process. We anticipate solving this problem very quickly once a set of target simulation environments is identified.
One next step includes application of this ASG process to a diverse set of task domains for both individual learners and teams. The application of ASG for teams is particularly challenging given the complexity of assessing team learning and performance, and will only strengthen the process over time.
Another next step for GA-based process might also include alignment with other learner/team states that moderate learning. For example, another desired end state for this research is to be able to adapt existing scenarios or select available scenarios based on individual learner emotions (e.g., boredom or anxiety) or behavioral markers (e.g., encouragement) which are antecedents to team states (e.g., team cohesion). The emotional state of the learner(s) might also be a criteria for selecting either a more or less difficult scenario or injecting support feedback (scaffolding) into an existing scenario to make it easier per Vygotsky's (1987) ZPD (Figure 3).
We also anticipate experimenting with local search after the crossover and mutation steps to see if that will yield better solutions or enhance the speed of the ASG process for more complex domains.
Finally, in spite of the complications in using deep reinforcement learning (Rowe, Smith, Pokorny, Mott, & Lester, 2018), we anticipate continuing research in this area with the hope of enhanced results over time. This assumes that we will find a process that will be flexible enough to apply easily in a variety of task domains trained by military organizations.

ACKNOWLEDGMENTS

The research described herein has been sponsored by the U.S. Army. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2006, June). The cognitive tutor authoring tools (CTAT): preliminary evaluation of efficiency gains. In *International Conference on Intelligent Tutoring Systems* (pp. 61-70). Springer, Berlin, Heidelberg.

Baker, James E. (1987). "Reducing Bias and Inefficiency in the Selection Algorithm". *Proceedings of the Second International Conference on Genetic Algorithms and their Application*. Hillsdale, New Jersey: L. Erlbaum Associates: 14–21.

Brawner, K. & Sottilare, R. (2018, June). Proposing Module-level Interoperability for Adaptive Instructional Systems. In the Exploring Opportunities to Standardize Adaptive Instructional Systems (AISs) Workshop of the *18th International Conference of Artificial Intelligence in Education (AIED)*, London, United Kingdom, June 2018.

Brown, D., Bell, B. and Goldberg, B., 2017. Authoring Adaptive Tutors for Simulations in Psychomotor Skills Domains. In Proceedings of *MODSIM World 2017*, Virginia Beach, VA: NTSA.

Dave, R.H. (1970). Psychomotor levels. In R.J. Armstrong (Ed.), *Developing and Writing Behavioral Objectives*. Tucson, Arizona: Educational Innovators Press.

Folsom-Kovarik, J.T. & Brawner, K. (2018, May). Automating Variation in Training Content for Domain-general Pedagogical Tailoring. In Proceedings of the 6th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium. *US Army Research Laboratory*, Orlando, FL.

Harrow, A. (1972) A Taxonomy of Psychomotor Domain: A Guide for Developing Behavioral Objectives. New York: David McKay.

Lehman, J. (2012). Evolution through the search for novelty. Dissertation: *University of Central Florida*.

Lehman, J., & Stanley, K.O. (2011). Novelty search and the problem with objectives Genetic Programming Theory and Practice IX (pp. 37-56): *Springer*.

Liapis, A., Yannakakis, G. N., & Togelius, J. (2015). Constrained novelty search: A study on game content generation. Evolutionary computation, 23(1), 101-129.

Luo, L., Yin, H., Cai, W., Zhong, J., & Lees, M. (2016). Design and evaluation of a data-driven scenario generation framework for game-based training. *IEEE Transactions on Computational Intelligence and AI in Games*.

Mallawaarachchi, V. (2017). Introduction to Genetic Algorithms – Including Example Code. *Towards Data Science*. Retrieved from: https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3.

Mitchell, Melanie (1996). An Introduction to Genetic Algorithms. Cambridge, MA: *MIT Press*. ISBN 9780585030944.

Mitrovic, A., Suraweera, P., Martin, B., Zakharov, K., Milik, N., & Holland, J. (2006, June). Authoring constraint-based tutors in ASPIRE. In *International Conference on Intelligent Tutoring Systems* (pp. 41-50). Springer, Berlin, Heidelberg.

Romiszowski, A. (1999). The development of physical skills: Instruction in the psychomotor domain. In Instructional-design theories and models: a new paradigm of instructional theory (Vol. 2). Mahwah, NJ: Erlbaum.

Rowe, J., Smith, A., Pokorny, R., Mott, B., and Lester, J., (2018, May). Toward Automated Scenario Generation with Deep Reinforcement Learning in GIFT. In Proceedings of the 6th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium. *US Army Research Laboratory*, Orlando, FL.

Shiffman, D. (2012). The Nature of Code: Simulating Natural Systems with Processing. Daniel Shiffman.

Simpson E.J. (1972). The Classification of Educational Objectives in the Psychomotor Domain. Washington, DC: Gryphon House.

Sinatra, A.M. & Sottilare, R. (2016). Chapter 14 – Exploring the Diversity of Domain Modeling for Training and Educational Applications. In R. Sottilare, A. Graesser, X. Hu, A. Olney, B. Nye & A. Sinatra (Eds.) Design Recommendations for Intelligent Tutoring Systems: Volume 4 – Domain Modeling. *US Army Research Laboratory*, Orlando, Florida. ISBN: 978-0-9893923-9-6 (digital version).

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The generalized intelligent framework for tutoring (GIFT). Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED).

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT). GIFTtutoring. org.

Sottilare, R. & Brawner, K. (2018a, March). Exploring Standardization Opportunities by Examining Interaction between Common Adaptive Instructional System Components. In Proceedings of the *First Adaptive Instructional Systems (AIS) Standards Workshop*, Orlando, Florida.

Sottilare, R. & Brawner, K. (2018b, June). Component Interaction within the Generalized Intelligent Framework for Tutoring (GIFT) as a Model for Adaptive Instructional System Standards. In the Adaptive Instructional System (AIS) Standards Workshop of the *14th International Conference of Intelligent Tutoring Systems (ITS)*, Montreal, Quebec, Canada, June 2018.

Sottilare, R.A. & Ososky, S. (2017, July). Defining Complexity in the Authoring Process for Adaptive Instruction. In *Foundations of Augmented Cognition* (pp. 237-249). Springer International Publishing.

US Army (2007). Field Manual No. 3-21.8 - The Infantry Rifle Platoon and Squad. Department of the Army Washington, DC, 28 March 2007

Vygotsky, L. (1987). Zone of proximal development. *Mind in society: The development of higher psychological processes*, 5291, 157.
Wikipedia. (2018). Brute-Force Search. Retrieved from: https://en.wikipedia.org/wiki/Brute-force_search
Zook, A., Lee-Urban, S., Riedl, M.O., Holden, H.K., Sottilare, R.A., & Brawner, K.W. (2012, May 29-June 1). Automated scenario generation: Toward Tailored and Optimized Military Training in Virtual Environments. Paper presented at the *International conference on the Foundations of Digital Games*, Raleigh, NC.

AUTHOR BIOGRAPHY

Dr. Robert A. Sottilare is the Adaptive Training Research Lead at the US Army Natick Soldier Center where the focus of his research is automated authoring, instructional management, and analysis tools and methods for intelligent tutoring systems (ITSs). He is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT), an open source, AI-based adaptive instructional architecture. He is also the lead editor for the Design Recommendations for Intelligent Tutoring Systems book series and founder/chair of the GIFT Users Symposia. Dr. Sottilare is a member of the IEEE and AI in Education Society. He leads an IEEE working group for adaptive instructional system (AIS) standards. He is widely published and is a frequent speaker at the Defense & Homeland Security Simulation (DHSS) Workshop, the Florida AI Research Society, Augmented Cognition, and AI in Education conferences. Dr. Sottilare is also a faculty scholar and adjunct professor at the University of Central Florida where he teaches a graduate level course in ITS design. He is also a frequent lecturer at the United States Military Academy (USMA) where he teaches a senior level colloquium on ITS design. He is the recipient of two lifetime achievement awards in Modeling & Simulation: US Army RDECOM (2012; inaugural recipient) and National Training & Simulation Association (2015; Governor's Award – highest level).

MODELLING AND SIMULATION OF A FIRE DEPARTMENT'S RESPONSE TO EMERGENCY INCIDENTS

Adriano O. Solis[(a)], Jenaro Nosedal-Sánchez[(b)], Ali Asgary[(c)], Francesco Longo[(d)], Beatrice Zaccaro[(e)]

[(a),(c)] School of Administrative Studies, York University, Toronto, Ontario, Canada
[(b)] Facultad de Ingeniería, Universidad Autónoma del Estado de México Cerro de Coatepec, Ciudad Universitaria Toluca, México
[(d)] DIMEG, University of Calabria, Rende (CS), Italy
[(e)] DIME, University of Genoa, Genoa (GE), Italy

[(a)] asolis@yorku.ca, [(b)] jnosedal@conacyt.mx, [(c)] asgary@yorku.ca, [(d)] francesco.longo@unical.it, [(e)] zaccarobeatrice@gmail.com

ABSTRACT
After statistical analysis of the database of a fire department covering eight years of consecutive incident records from January 2009 to December 2016, we developed a modelling and simulation (M&S) approach that could be replicated for fire departments across Canada. Our M&S framework involved two different simulation models running on separate platforms: (i) an *Incident Generation Engine*, which simulates the 'arrival' of emergency incidents, and (ii) a *Response Simulation Model*. The first model is a discrete event simulation model using CPNTools 4.0, generating inputs for the second model, which is an agent-based simulation model developed using AnyLogic. We discuss the principal elements of the two simulation models, and report on findings from our simulation experiments.

Keywords: Fire department, emergency incidents, discrete event simulation, agent-based simulation

1. INTRODUCTION
According to the 2013 CSSP (Canadian Safety and Security Program) environmental scan, the construction of risk models or algorithms to predict locations or circumstances with a high likelihood of becoming an incendiary incident as well as predictive models that are able to better estimate actual or potential losses are highly needed (Greene 2014). The same report states that national-level reporting, data, statistics, metrics, as well as risk assessment related to infrastructure and building codes are considered high priority items by the Fire Community of Practice (CoP). The CSSP report also points to the need for research projects that "enable Canada's emergency managers, responders and security authorities through timely access to science and technology expertise in support of planning, operations and exercises and assessing the performance of technologies and processes in simulated and operational settings." Our research seeks to address these needs.

There are four widely identified phases in the disaster and emergency management cycle (Haghani and Afshar 2009):

- Mitigation,
- Preparedness,
- Response, and
- Recovery.

Mitigation is aimed at reducing vulnerability to impacts of disasters and emergencies such as injuries, death, or damage to property. It may involve changes in municipal building codes and zoning, as well as analyses and strengthening of public infrastructure. Preparedness involves planning how to respond in terms of business continuity and emergency management, education, outreach, and training, and setting up warning systems. Even as disaster and emergency response and recovery are ongoing concerns, significant efforts by government agencies, researchers, and other sectors are understandably geared towards addressing mitigation and preparedness.

Academics and enterprise solution providers have developed different approaches and tools for using descriptive and predictive analytics towards safer or smarter cities. Research projects and tools, however, have mainly been applied on crime, energy, public transportation or water distribution concerns. The "Smart City Initiative" (Ghosh et al. 2016), for example, calls for application of smart computing for awareness of existing assets, public concerns and needs, and for developing advanced analytics to uncover the innovative improvements and predict underutilized potentials to improve the policy making process.

Prior work in data-driven urban residential risk prediction at the regional or community level has been undertaken by Clare et al. (2012) and DaCosta, Krinsley, and Abelson (2015), and at building level for the city of Atlanta by Madaio et al. (2016). Wuschke, Clare, and Garis (2013) applied a temporal and geographic clustering of crime events (e.g., shootings) in residential areas for developing a crime prevention tool. They analysed the patterns of residential burglary and residential structure fire events across time and space within a large North American city. Guidelines for the use of various methods in detecting the differences in clustering patterns of fire and non-fire incidents for urban data are provided by Ceyhan, Ertuğay, and Düzgün (2013).

An example of earlier research applying probabilistic simulation of fire incidents can be found in Hostikka and Keski-Rahkonen (2003). However, applicability is limited due to the lack of empirical data to determine the distributions, and only two restrictive applications, to (i) a cable tunnel fire and (ii) an electronics cabinet fire in an electronics room, are presented. Another example of stochastic modeling is the work of Chu and Wang (2012) who presented an approach to analyze probability distributions of fire scenarios in risk assessment for emergency evacuation. Salem (2016) reported the application of micro-modeling and Monte Carlo simulation to predict the consequences of fires in roll on-roll off/passenger ships and cruise liners, particularly in terms of available safe egress times which are critical to the safety of passengers and crew members. Similarly, Kang et al. (2017) report fire simulations applied during the early phases of the ship design process.

With respect to fire department operations, Halpern, Isherwood, and Wand (1979) analysed the relationship for residential fires between fire department response times and fire losses. The analysis was based on actual single and double family dwelling fires which occurred in the city of Calgary, in the Western Canadian province of Alberta, in 1976 and 1977. The study was intended to enable decision makers to consider the benefits of reducing response times by way of additional firefighting resources (fire stations, vehicles and other equipment, and personnel).

Carroll (1982) developed a simulator consisting of two programs:

- the first program generating – based on historical data (actual fire history for the three years 1976, 1977, and 1978 of the city of London, in the Canadian province of Ontario, as handwritten in a bound ledger called the Record of Fires, Alarms, and Operations) – a chronological file of simulated fire occurrences over ten years; and
- the second program processing the simulated events to produce statistics with respect to availability and utilization of firefighting resources under various configurations.

The simulator was intended to help the city decide how many fire stations are needed and where they should be located.

More recently, Aleisa and Savsar (2013) applied discrete event simulation to firefighting operations in Kuwait, to test various changes in fire department operating policies. The objective was to assess whether a change, or set of changes, would lead to response times of under five minutes in all districts. Vaira, Hammond, and Rusnock (2016) developed a discrete event simulation model of turnout times at a single fire station, using task times derived from a time and motion study fitted to probability distributions. Testing a procedural/behavioral change in their case study resulted in simulated turnout times decreasing by 24.3%.

Altay and Green (2006) surveyed 109 papers on disaster operations management (DOM) published over the 25 years from 1980 and 2004 in mainstream operations research/management science (OR/MS) journals. They found that 11.9% of the papers involved modelling and simulation (M&S) as methodology. Galindo and Batta (2013) undertook a subsequent survey over the six-year period from 2005 to 2010, and they found that 9% of the 155 papers surveyed used simulation. While mathematical programming was reported to be the most preferred methodology (32.1% in the earlier survey; 23.1% in the latter), simulation is cited to have maintained its status as a widely used OR/MS methodology in DOM research (Galindo and Batta, 2013).

2. DATASET AND ANALYSIS

We had sought to undertake an extensive analysis based on the unique opportunity to investigate 11 years (2005 through 2015) of comprehensive fire incident records in the National Fire Information Database (NFID) for Canada, and build reliable prediction models for the most relevant factors that describe a fire incident (Solis et al. 2018). The NFID is made available in two main data files, one with information about incidents and the other containing information about victims (civilians and firefighters) for the 11 years. The NFID as of July 2017 contained 136 fields, which correspond to specific attributes that should be recorded/reported for each incident, all defined/explained, along with the coding values (i.e., individual categories, counts, magnitudes, etc.) in an *NFID Data Dictionary* and an *NFID User Guide*.

While the NFID represents a set of relevant data for the analysis of various factors associated with the occurrence of fires, a report issued by the the Canadian Centre for Justice Statistics (2017), prepared for the Canadian Association of Fire Chiefs, indicated very serious data gaps. In our own earlier macro-analysis of NFID incident and victim datasets, we detected a very significant amount of missing values (blanks) that would have constituted critical inputs for our purposes. For example, a very critical data field for M&S of fire department operational performance is *RESPONSE TIME*, which is defined in the NFID Data Dictionary as "the period of time from the receipt of the alarm by the Fire Department to the arrival of the first vehicle at the scene of the incident." Unfortunately, response time information is quite scarce in the NFID. Furthermore, the NFID is only limited to fires and fire-related incidents. We have found that such incidents do not even constitute the majority of incidents that fire departments respond to. In Figure 1, for instance, we see that emergency fire incidents accounted for only 28.1% of incidents that Toronto Fire Services responded to in 2016. The City of Toronto is the largest in Canada in terms of population, with the number of inhabitants pushing close to three million.

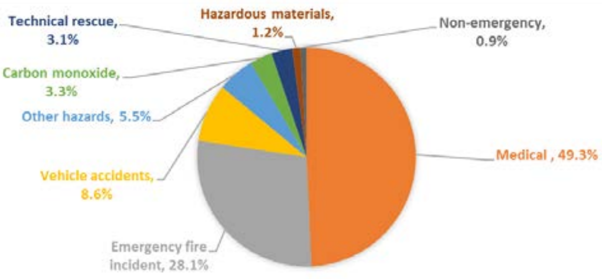

Figure 1: Toronto Fire Services – Breakdown of 2016 Emergency Incidents (Source: Solis et al., 2018)

Similarly, for the city of Calgary, which is the third largest Canadian city with a population of just over one million, the Calgary Fire Department reported that 26% of emergencies it responded to in 2016 were fire and fire related incidents (Figure 2).

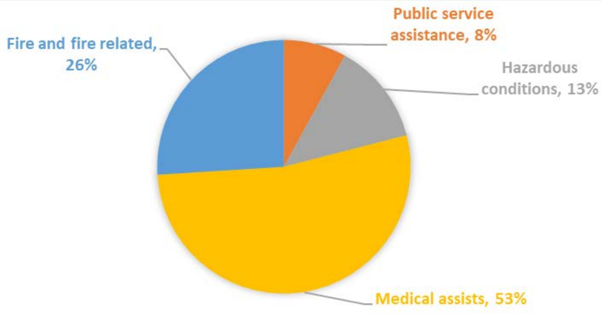

Figure 2: Calgary Fire Department – Breakdown of 2016 Emergency Incidents (Source: Solis et al., 2018)

In light of the serious gaps we observed in key fire department operational data (incident location, time of alarm receipt, response time, etc.) in the NFID, coupled with the fact that the NFID is limited only to fire and fire related incidents, we sought the assistance of the Vaughan Fire and Rescue Service (VFRS). The city of Vaughan, which is located just north of Toronto, is the 17th largest Canadian city with a population of more than 300,000. VFRS (2016) reported that property fires and explosions accounted for 9% of emergency incidents in 2016, with an additional 16% in false alarm calls (see Figure 3). The remaining 75% of incidents were principally from medical emergencies (41%), rescues (18%), and others (16%).
Our aim was to develop a fairly generic model that could be replicated for fire departments across Canada, for as long as the appropriate set of operational data are collected and maintained by such other fire departments. The VFRS data cover eight years of consecutive incident records from January 2009 through December 2016. Because the VFRS incident dataset included all types of emergencies (see Figure 2) – not just fires and fire related incidents, as in the case of NFID – it became apparent that a better understanding and analysis of incident occurrences, fire department responses, and resulting requirements for, and utilization of, fire department resources (firefighting vehicles and crews) was attainable.
We decided to develop the simulation framework taking into account the standard information available at fire department level in the case of VFRS (in accordance with directives issued by the Office of the Fire Marshal of Ontario).

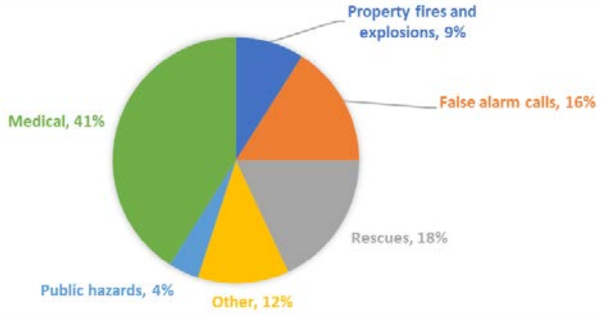

Figure 3: VFRS – Breakdown of 2016 Emergency Incidents (Source: Solis et al., 2018)

It is possible to extract from the VFRS data a set of key features related with incidents and response characteristics. Some fields coincide with those reported in the NFID on the incident information fields, such as *INCIDENT ID*, *ALARM TIME*, *RESPONSE TIME* and *INCIDENT LOCATION*. Some relevant fields in addition to those in NFID are: *TYPE OF INCIDENT* and *ON SCENE TIME*.
We initially undertook an assessment of data availability and quality. We sought to eliminate wrongly coded values and outliers. Of the above mentioned key data fields, the worst case – for *ON SCENE TIME* – provided 88% of utilizable data (available records after the cleanup process).
From our analysis of time intervals (in minutes) between consecutive values of *alarm time* (or incident “inter-arrival times”), we obtained the inter-arrival distribution depicted in Figure 4. It suggests a negative exponential function, which is consistent with a Poisson distribution that commonly characterizes the arrival of customers into a service queuing system.
Figure 5 shows the overall distribution of *RESPONSE TIME* in the VFRS dataset. However, we analysed the response times observed for each type of incident, obtaining different mean and variability statistics. In our simulation studies, we used the actual empirical distributions of response times according to various incident types. A similar analysis of *ON SCENE TIME* showed different shapes and parameters across various incident types.

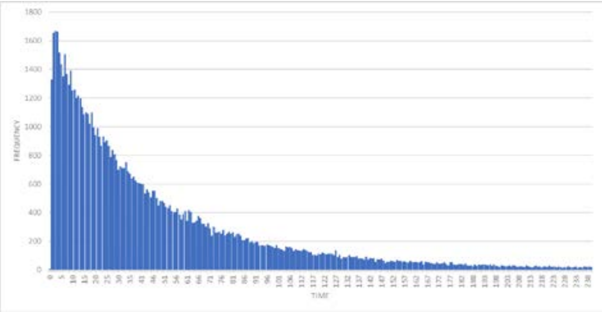

Figure 4: VFRS Data – Distribution of Time between Successive Calls (Source: Solis et al., 2018)

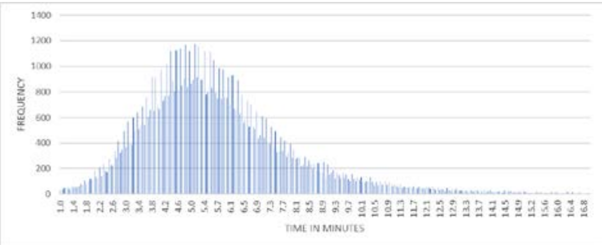

Figure 5: VFRS Data – Distribution of Response Times (Source: Solis et al., 2018)

The VFRS dataset includes as data fields the Alarm Receipt Time stamp (when the alarm is received at the VFRS communications centre) and the Roll-out Time stamp (when the vehicle rolls out of its home station). We were accordingly able to obtain, for each reported incident, an *Alarm Processing Plus TurnOut Time (APPTOT)*, which is calculated as Roll-out Time stamp minus Alarm Receipt Time stamp. We note that the two APPTOT component times – (i) Alarm Processing Time and (ii) Turnout Time – are not separately recorded in the VFRS dataset (which complies with Office of the Fire Marshal of Ontario directives). We used the Roll-out Time stamp of the first responding unit at the scene in calculating the APPTOT. Since the originating station is not clearly identified by way of a separate data field, we assumed that the first responding unit originates from the station responsible for the given emergency point. The distribution of *APPTOT* is presented in Figure 6.

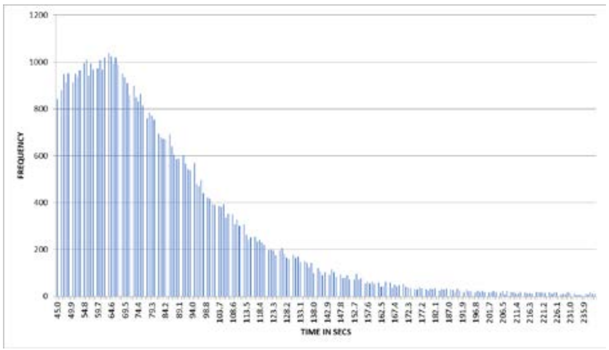

Figure 6: VFRS Data – Distribution of APPTOT (Source: Solis et al., 2018)

3. MODEL DESCRIPTION

We developed an M&S framework involving two different simulation models running on separate platforms:

1. An Incident Generation Engine simulating the ‘arrival’ of incidents, which we developed and implemented as a discrete event simulation model using CPNTools 4.0 (CPNTools 2017). Each emergency incident generated is characterized by:
   - incident ‘arrival’ time (the Alarm Receipt Time stamp at the VFRS communications centre),
   - incident type,
   - incident location (GIS coordinates),
   - APPTOT, and
   - on scene time.
2. A Response Simulation Model, which we developed and implemented using agent based modelling within AnyLogic 8.0.5 (AnyLogic 2017).

3.1. Incident Generation Engine

The Incident Generation Engine has been developed and implemented as a discrete event simulation model with CPNTools 4.0 as the platform. It produces a list of emergency incident occurrences (including the key incident features: incident ‘arrival’ time, incident type, incident location, APPTOT, and on scene time), generated based on the empirical distributions for the key incident features. This chronological list of incidents provides the inputs for the Response Simulation Model.

The discrete event model simultaneously takes information from the empirical distributions, considering central tendency and variability statistics, of various data fields within the VFRS dataset, and generates the following for each incident occurrence:

1. incident ‘arrival’ time,
2. incident type,
3. incident location,
4. APPTOT, and
5. on scene time.

Incident ‘arrival’ time is based on the distribution of inter-arrival times between successive alarms received. Incident type is identified considering the probabilities of occurrence of specific types of incidents over time during the day. We created a partition of the entire geographical region covered by VFRS, using a lattice granularity of 500 m × 500 m (0.25 km^2). The incident location is based on the 500 m × 500 m cells in the lattice. APPTOT and on scene time are set for the first responding firefighting vehicle/crew. The Incident Generation Engine output list includes, for each incident, an incident ID plus the five data fields specified above.

Figure 7 provides an overview of the main components of the model (places, transitions, and arcs) for the Incident Generation Engine. Descriptions of the components (network nodes) are summarized in Table 1.

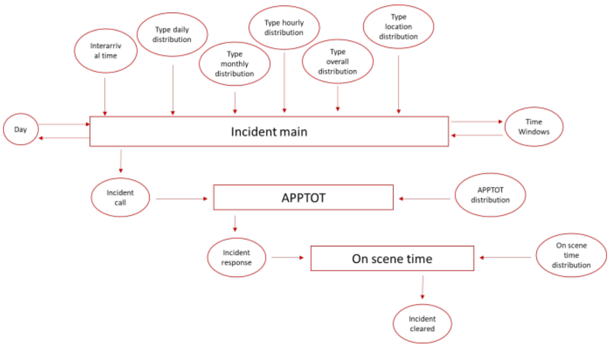

Figure 7: Main Components of the Incident Generation Engine

3.2. Response Simulation Model

The incident list created using the Incident Generation Engine serves as input into the second model – the Response Simulation Model. While the Incident Generation Engine is a discrete event simulation model, we used Agent-Based Modelling (ABM) to develop our Response Simulation Model, using the AnyLogic 8.0.5 simulation platform (AnyLogic 2017). ABM is best described as a decentralized, individual-centric approach to modelling. In this approach, individual participants have their own behavior and are referred to as agents.

The agents that come into play in the model are:

- Dispatcher – the person at the VFRS communications centre who receives the emergency call and alerts the appropriate Station to respond to the incident;
- Emergency Point – an entity that represents the location of an emergency, which changes its status based on the actions of the other agents;
- Station – the agent that receives the call from the Dispatcher and changes its status according to availability of resources; and
- Vehicle – the entity that receives the dispatch order from the available Station. It uses the GIS Map to determine and use the appropriate route.

Based on VFRS practice, a crew of four full-time firefighters mans each responding Vehicle. There is at least one spare fire engine at every Station to ensure availability.

Within the AnyLogic simulation platform, we are able to transform reported longitude and latitude information on Emergency Points as well as Stations into usable Geographic Information System (GIS) coordinates which allow simulation of vehicular travel on city streets.

Table 1: Descriptions of Network Nodes of the Incident Generation Engine

Node Name	**Node Description**
Transition Nodes	
Incident main	Takes one value from each of the places (nodes) connected to simulate the main incident features (inter-arrival time, incident type, incident location).
APPTOT	Takes one value from the distribution depending on incident type.
On scene time	Takes one value from the distribution depending on incident type.
Place nodes	
Day	Contains information on day (Monday, Tuesday, etc.) and month (January, February, etc.) for generation of the incident list.
Inter-arrival time	Contains the inter-arrival time, obtained from the distribution of times between successive calls.
Type overall distribution	Contains the incident occurrence distribution per location, for each incident type.
Type monthly distribution	Contains the incident occurrence distribution per month, for each incident type.
Type daily distribution	Contains the incident occurrence distribution per day, for each incident type.
Type hourly distribution	Contains the incident occurrence distribution per hour, for each incident type.
Type location distribution	Contains the incident occurrence distribution per location (cells of 500 m x 500 m), for each incident type.
Time windows	Contains the incident occurrence distribution within each of four six hour time windows, for each incident type.
Incident call	Contains the incidents generated along with the main features (inter-arrival time, type, location).
APPTOT distribution	Contains the estimated APPTOT, obtained from the distribution for overall incident types.
Incident response	Contains the incidents generated along with the main features, plus the preparation time.
On scene time distribution	Contains the incident on scene time distribution, for each incident type.
Incident cleared	Final node with all features simulated: type, location, inter-arrival time, APPTOT, on scene time.

Animation introduced into a scaled representation of the city's five Fire Districts and nine Stations helps the user to visualize what is happening in the system in terms of entities and the flow of resources (see Figure 8). Table 2 shows the number of responding units (vehicles/crews) available at each Station.

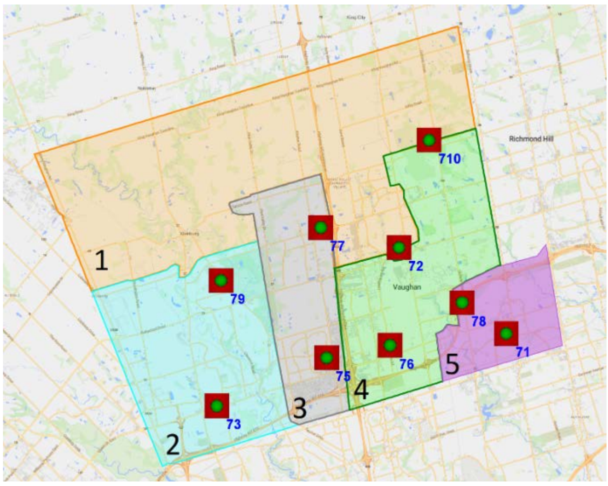

Figure 8: VFRS Five Fire Districts and Nine Fire Station Locations

Table 2: VFRS Current Number of Responding Units per Station

Station Code	Fire District	No. of Vehicles/Crews
71	5	2
72	1	2
73	2	2
75	3	2
76	4	1
77	3	1
78	5	1
79	2	1
710	1	1

The Incident Generation Engine output list, as earlier specified, would translate into inputs to the Response Simulation Model are as follows:

1. Incident ID,
2. Incident arrival time,
3. Incident type (Emergency Code),
4. Latitude,
5. Longitude,
6. APPTOT,
7. On scene time, and
8. Number of responding vehicles needed.

In particular, VFRS follows a protocol for the number of responding vehicles depending upon the emergency code – 1, 2, or 4 vehicles. In applying this protocol, we obtained, based on the VFRS historical data, the distribution of number of responding units required as shown in Table 3. We see a relatively small percentage (5.3%) of incidents requiring more than two responding units. Accordingly, we decided to use either one or two responding units depending on the emergency code, with any code requiring four responding units being assigned two units instead. This enables a relatively simple simulation model, but with simulation results still roughly indicative of resource requirements.

Table 3: VFRS Expected Distribution of Number of Responding Vehicles per Current Protocol

No. of Responding Vehicles per Protocol	% of Incidents
1	60.5%
2	34.1%
4	5.3%
Total	100.0%

4. PRELIMINARY RESULTS

A replication in our experiment simulates one day of VFRS responses to emergency incidents. We conducted and collected statistics from 365 replications (corresponding to a one-year period from January 1 to December 31) of the experiment. In effect, 365 days of incidents were generated using the *Incident Generation Engine*, and the resulting incident list was used as input for the *Response Simulation Model*.

Figure 9 shows the overall percentages of simulated response plus on scene times across all incidents throughout the one-year period, broken down into APPTOT (5%), travel time (40%), and on scene time (55%).

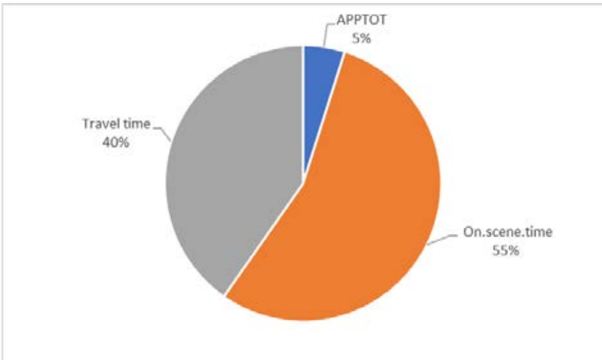

Figure 9: Aggregate Percentages of Simulated Response Times and On Scene Times

When reported according to fire districts, however, the distributions differ across regions (see Figure 10). For instance, travel times account for more than 50% of total time in the case of Fire District 1, which includes Stations 71 and 710. A question to consider is whether it would be appropriate to relocate either or both Fire District 1 stations.

One of the statistics of interest is the utilization of responding units at each Station. For instance, Station 76 (in Fire District 4) is presently allocated one responding unit (a fire engine with a four-person crew).

In our 365-day simulation experiment, Station 76 has no responding unit remaining 11.2% of the time (see Figure 11). A relevant question might be whether a second responding unit should be considered.

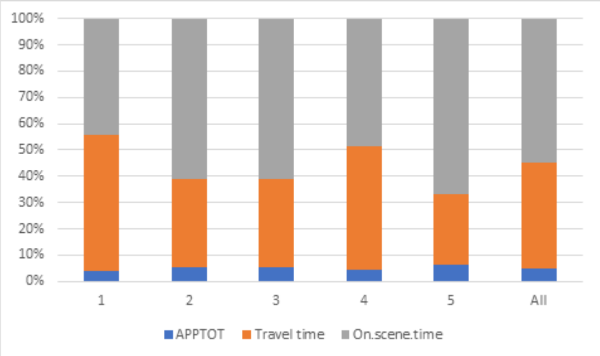

Figure 10: Overall Distribution of Simulated Times, by Fire District

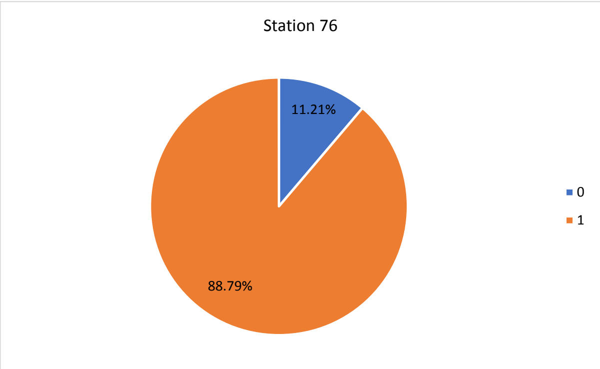

Figure 11: VFRS Station 76 – Simulated Responding Unit Utilization

5. CONCLUSION

Based on the preliminary simulation results, which will be further reported on at the time of the conference, questions with respect to current VFRS operating parameters – e.g., location of fire stations, number of responding units (firefighting vehicles/crews) per fire station – and protocols arise. We are also in the process of refining our Response Simulation Model to better address and seek answers to such questions.

ACKNOWLEDGMENTS

This research project is conducted under the umbrella of the Advanced Disaster, Emergency and Rapid Response Simulation (ADERSIM) program at York University. The Modeling and Simulation Center–Laboratory of Enterprise Solutions (MSC-LES) of the University of Calabria, Italy, is an international collaborator of the ADERSIM program.

The work has been supported by a grant from the Canadian Association of Fire Chiefs (CAFC) in support of a research project on 'Developing a fire response test bench based on NFID'. The project had sought to develop a simulation engine leveraging the National Fire Information Database (NFID).

We also wish to acknowledge partial financial support from Dr. Jianhong Wu, University Distinguished Research Professor and Canada Research Chair in Industrial and Applied Mathematics, York University.

We are grateful to recently retired Chief Larry Bentley, current Chief Deryn Rizzi, Deputy Chief Andrew Zvanitajs, and Captain Kevin Plested of the Vaughan Fire and Rescue Service, not only for making available the VFRS incident dataset, but most especially for providing assistance, guidance, and insights in our statistical analysis of the VFRS dataset, ongoing evaluation of preliminary simulation results, and further refinement of our simulation models.

REFERENCES

Aleisa, E., and Savsar, M., 2013. Modeling of firefighting operations through discrete event simulation. International Journal of Computer Theory and Engineering, 5 (4), 678-682.

Altay, N., and Green, W.G., 2006. OR/MS research in disaster operations management. European Journal of Operational Research, 175 (1), 475-493.

AnyLogic, 2017. AnyLogic Simulation Software Version 8.0.5. Accessed in October 2017 at http://www.anylogic.com.

Canadian Centre for Justice Statistics, 2017. Fire Statistics in Canada, Selected Observations from the National Fire Information Database 2005 to 2014. Report prepared for the Canadian Association of Fire Chiefs, September 2017.

Carroll, J.M., 1982. Digital simulator for allocating firefighting resources. Simulation, 38 (1), 1-12.

Ceyhan, E., Ertuğay, K., and Düzgün, S., 2013. Exploratory and inferential methods for spatio-temporal analysis of residential fire clustering in urban areas. Fire Safety Journal, 58 (complete), 226-239.

Chu, G., and Wang, J., 2012. Study on probability distribution of fire scenarios in risk assessment to emergency evacuation. Reliability Engineering & System Safety, 99 (complete), 24-32.

Clare, J., Garis, L., Plecas, D., and Jennings, C., 2012. Reduced frequency and severity of residential fires following delivery of fire prevention education by on-duty fire fighters: Cluster randomized controlled study. Journal of Safety Research, 43 (2), 123-128.

CPNTools, 2017. CPNTools 4.0. Downloaded from http://cpntools.org/.

DaCosta, M., Krinsley, J., and Abelson, B., 2015. Optimizing local smoke alarm inspections with federal data. Bloomberg Data for Good Exchange 2015.

Galindo, G., and Batta, R., 2013. Review of recent developments in OR/MS research in disaster operations management. European Journal of Operational Research, 230 (2), 201-211.

Ghosh, D., Chun, S.A., Shafiq, B. and Adam, N.R., 2016. Big data-based smart city platform: Real-time crime analysis. Proceedings of the 17th

International Digital Government Research Conference on Digital Government Research, pp. 58-66. June 8-10, Shanghai, China.

Greene, B.W., 2014. CSSP environmental scan 2013. Scientific Report DRDC-RDDC-2014-R83, Defence Research and Development Canada, September 2014.

Haghani, A., and Afshar, A.M., 2009. Supply chain management in disaster response. Final project report submitted to the Mid-Atlantic Universities Transportation Center, September 2009. Accessed at https://rosap.ntl.bts.gov/view/dot/34430.

Halpern, J., Isherwood, G.D., and Wand, Y., 1979. The relationship between response time and fire property losses. Infor, 17 (4), 373-389.

Hostikka, S., and Keski-Rahkonen, O., 2003. Probabilistic simulation of fire scenarios. Nuclear Engineering and Design, 224 (3), 301-311.

Kang, H.J., Choi, J., Lee, D., and Park, B.J., 2017. A framework for using computational fire simulations in the early phases of ship design. Ocean Engineering, 129 (complete), 335-342.

Madaio, M., Chen, S-T., Haimson, O.L., Zhang, W., Cheng, X., Hinds-Aldrich, M., Chau, D.H., and Dilkina, B., 2016. Firebird: predicting fire risk and prioritizing fire inspections in Atlanta. http://www.kdd.org/kdd2016/papers/files/adf0511-madaioA.pdf.

Salem, A.M., 2016. Use of Monte Carlo Simulation to assess uncertainties in fire consequence calculation. Ocean Engineering, 117 (complete), 411-430.

Solis, A.O., Asgary A., Nosedal-Sánchez, J., and Zaccaro, B., 2018. Developing a fire response test bench based on NFID. Project report submitted to the Canadian Association of Fire Chiefs, 1 March 2018.

Vaira, K., Hammond, G., and Rusnock, C., 2016. Improving fire station turn-out-time using discrete-event simulation. Proceedings of the 2016 Winter Simulation Conference, pp. 3686-3687. December 11-14, Washington, DC, USA.

VFRS, 2016. Vaughan Fire & Rescue Service 2016 Annual Report.

Wuschke, K., Clare, J., and Garis, L., 2013. Temporal and geographic clustering of residential structure fires: A theoretical platform for targeted fire prevention. Fire Safety Journal, 62 (Part A), 3-12.

SIMULATION OF CRISIS AFFECTING CRITICAL INFRASTRUCTURES AND INDUSTRIAL PLANTS

Agostino G. Bruzzone[(a)], **Marina Massei**[(b)], **Riccardo di Matteo**[(c)], **Matteo Agresta**[(d)]

[(a), (b)] DIME University of Genoa,
[(c)] SIM4Future - [(d)] Simulation Team

[(a), (b)] { agostino, massei, agresta}@itim.unige.it
[(c)] riccardo.dimatteo@sim4future.com - [(d)] matteo.agresta@simulationteam.com

[(a), (b)] www.itim.unige.it - [(c)] www.sim4future.com - [(d)] www.simulationteam.com

ABSTRACT

This paper proposes a model for reproducing human behaviors and it is devoted to evaluate the operation performance in case of crisis with special attention to cases involving multiple actors such as fire-fighters, police, ambulances, civil protection and national reserve. The paper present the general scheme and the data structure used to conduct experiments related to cases of flooding and hazardous material spills in industrial and urban areas.

Keywords: Crisis Simulation, Interoperable Simulation, Industrial Crisis, Decision Support

1 INTRODUCTION

Today urban areas have growth up to incorporate many industrial facilities; due to this fact population density near to industrial plants dealing with hazardous material increased drastically, as well as concreting and land use; these factors increase the risks for population to be involved in crisis due to natural and/or man-made disasters (Mengfan 2016; Harding J.). It is evident that the urban framework introduces additional challenges for first responders in order to operate in the area and coordination issues turn to be even more complicated than in usual operations. These phenomena are highly stochastic, affected by multiple interactions of many entities over different layers as well as strongly influenced by external (e.g. weather conditions) and human factors; so it is evident the necessity to use Modeling and Simulation to model the different elements of the real systems (Longo 2013) and anticipate their behaviors (Longo et al. 2008). In this paper the authors propose part of a larger model in use to simulate crisis, able to reproduce industrial processes and phenomena ingenerating the crisis evolution as well as population behavior and first responder actions.

2 CRISES IN INDUSTRIES AND CRITICAL INFRASTRUCTURES

The authors have broad experience in dealing with simulation of a crisis including explosions, hazardous material spills, fires, flooding, terrorist attacks, etc. (Bruzzone et al., 2008, 2014, 2016b).

In facts during these crises, multiple actors are present and they need to be coordinated: fire fighters, police, ambulance & health care facilities, national reserve, civil protection, etc.

This complex organizations could adopt different models for operating together and it is important to define the most effective and be aware of all the benefits from different structures, policies, regulations, operative plans and even technologies supporting these aspects. For instance, a major issue in a crisis dealing with an industry or a critical infrastructure is related to his impact on people; this means that transportation network could be compromised, often also power distribution is an issue to be solved as well as other urban layers. Moreover, gas distribution network in case of collapse of a critical infrastructure need to be secured in order to avoid further disasters or explosions. In this framework one of the major issue is related to people evacuation from dangerous areas (e.g. flooded or contaminated): this represent a major effort that could lead easily with hundreds of families or thousands of people to be moved within a town during chaotic moments. Another crucial aspect is related to the time scale that requires to address not only first response, but also several actions on short terms, required by the actors that deal with public safety and security.

The authors have currently specific simulation models and libraries for addressing these aspects including IA-CGF, ST_CIPROS, ST_CRISOM, PONTUS, etc. (Bruzzone et al.,2014, 2016a, 2017a). These simulators adopt interoperability through High Level Architecture Standard in order to guarantee the possibility to federate together different models and distributed with other systems (Kuhl et al.2000); in this way it is possible to substitute the hazardous material spill model for real-time simulation devoted to training, with high fidelity simulators devoted to reproduce the contamination on the area for analyzing specific scenarios; at the same time, reliable models for fast time simulation could be used to run multiple replications in order to finalize in a preventive way the Standing Operation Procedures (SOP) respect specific period of the year (Bruzzone et al. 1996, 2017c). Interoperability is fundamental, it is useful also for evaluating the coordination respect different actors; indeed by this approach different simulators could be federated allowing each of the actor to play a scenario with others and to be trained on their Command and Control (C2) systems; from this point of view there are several researches for defining capabilities, new guidelines and standards to reinforce use of simulation within different major actors during a crisis.

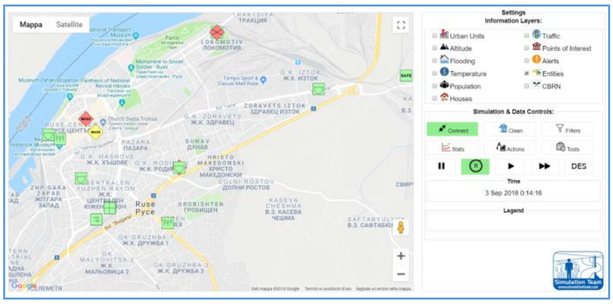

Fig.1 ST_CRISOM Simulation

The authors have recently completed different experimentations on projects dealing with crisis. First mention is about Decision Theater where different simulation models have been used such as ST_CRISOM, CIPROS, PONTUS within the Smart City funding program; the main crisis event was related to flooding in urban area with the possibility to diffuse contamination by depleting chemical tanks and facilities in industrial areas. In this case, the simulator addressed different criticalities:

- emergency management and evacuation considering the impact of flooding and contaminated waters,
- strategic planning in the given area to secure the situation by investments in infrastructures,
- industrial facilities additional safety elements,
- preventive crisis plans for first responders.

It is evident that this approach support also safety engineering and analysis devoted to prevent accidents and mitigate effect of crises. The experimentation allowed also to test decision makers' needs and expectations together with the City Authority and it was applied to Genoa case (major Italian port Town with less than a million inhabitants). The research tested also the possibility to interconnect the simulator with social networks to perform on-line acquisition of information coming directly from people (e.g. manhole blockage, river level, traffic jam, visible contamination, water level, etc.); the developed model fuses these information with other available data and spreads the input over time and space considering the constraints (Bruzzone et al.2017b).

Another application case is currently related to researches about use of these models for supporting collaboration between military units and civil agencies during such crises (Bruzzone et al. 2017a).

Also in these case ST_CRISOM, CRISOM, PONTUS, CIPROS of Simulation Team were tested and federated with other systems (i.e. jcats, pac, cora, hpac, dm, sword, arcaria-sword) by using different FOM popular among defense applications. In this case population has owned and controlled by Simulation Team as well as some specific crisis (e.g. flooding and industrial contamination or hazardous material spills).

Currently the authors are working in using these simulators in application to industrial cases to evaluate the impact of hazardous material spills and corresponding preventive actions and emergency management. The recent collapse of the "Morandi Bridge", a tragic event affecting Genoa (Harding 2018), makes the ST_CRISOM model, already tuned for flooding from previous Smartcity project, an additional way to provide support for mitigating the crisis and possibly restoring the affected areas. The bridge was an actual critical infrastructure, quite large, connecting Genova with its port. By the way, this specific area of the town statistically collected the highest rainfall in whole Italy (Acquaotto et al.2017; Faccini et al.2018). In addition the bridge was crossing a very high density populated district and the major industrial zone with major facilities (e.g. Ansaldo Energy). This case introduces the destruction of a critical infrastructure that impacts the Port Terminals, the transportations over the region (East to West and vice versa) and in some way the whole mobility.

Over 600 people evacuated and large industrial complex need to be reorganized and relocated. In case of heavy rain falls, the ruins on the river could potentially increase flooding risk as well potential contamination of waters/mud, being the fall season the most dangerous period on this area. Inhabitants' high density, heavy concreting and soil consumption affect the simulation evolution through water speed as well as the impact on the river bed and channels created by urbanization (Duley 1939; Chanson 2004). Currently the authors are working on this case to evaluate alternative solution for transportations, alternative actions to address different risks taking care of preventive and mitigation effects; this paper presents some of the data structures used to model this case, especially ST_CRISOM simulator (Simulation Team Crisis Simulation, Organization and Management) that reproduces dynamics of a given complex scenario where a crisis evolves.

3 CRISIS SIMULATION

ST_CRISOM considers the human behavior of the population in terms of evacuations, reactions due to the emergency as well as to human factors (fear, stress, fatigue and aggressiveness). ST_CRISOM is a stochastic hybrid interoperable simulator using intelligent agents; indeed ST_CRISOM combines discrete event simulation of operations with continuous simulations of physical systems such as flooding.. Indeed ST_CRISOM uses the IA-CGF (Intelligent Agent Computer Generated Forces) to reproduce both civilian Populations as well as First Responders and Military units, Health Care, Civil Protection Agents & Public Infrastructures (Bruzzone 2013). CRISOM acts as a NCF (Non Conventional Framework) for IA-CGF and allows to reproduce Flooding Scenario and Industrial Contamination as well as Hazardous Material Spill over regional areas and impact on Town, Industrial Facilities and Critical Infrastructures. ST_CRISOM adopts HLA (e.g. evolved HLA IEEE1516-2010) standard to be federated with other Simulators using different RTIs.(e.g. Pitch RTI 5.3.2.1).

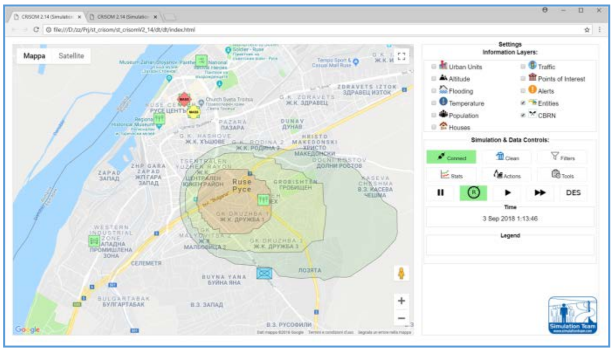

Fig.2 ST_CRISOM Simulation

4 PLAYERMESSAGE

ST_CRISOM uses a specific Interaction, defined PlayerMessage, for managing communications, lists and reports with other tools or simulators that have constraints is fully access HLA features.

Player Message in the FOM

```
<interactionClass>
    <name>PlayerMessage</name>
    <sharing>PublishSubscribe</sharing>
<transportation>HLAreliable</transportation>
    <order>Receive</order>
    <semantics>CIPROS player message.</semantics>
    <parameter>
      <name>Time</name>
      <dataType>HLAfloat64Time</dataType>
      <semantics>
Time when the report has to be sent (if the time is not specified the report has to be delivered as soon as possible) or an order/instruction has to be executed (if the time is not specified the order has to be executed as soon as possible).
Please note: The same info proposed here is available in the PlayerMessage content!</semantics>
    </parameter>
    <parameter>
      <name>Source</name>
<dataType>HLAunicodeString</dataType>
      <semantics>The Player who has to execute the order or the sender of the report (Player ID).
Please note: The same info is available in the PlayerMessage content!</semantics>
    </parameter>
    <parameter>
      <name>Content</name>
<dataType>HLAunicodeString</dataType>
      <semantics>the content of the message. (JSON format)</semantics>
    </parameter>

</interactionClass>
```

Player Message Examples

Examples of Format used by ST_CRISOM through Player Message in order to transmit to other systems specific info:

Code 36 Status Update of Critical Infrastructure
Update on the status of an industrial plant, tank or critical infrastructure during the crisis

[report id][unit id][time][location][Status: 0-100%][B]

[order id], *Status Update, [unit_id], [time], [location], [Status: 0-100%] , [free text] $

Code 37 Weather Update
Update on weather situation

[report id][type][pdi_id][time][location][wind speed][wind direction][rain intensity mm/h][snow intensity mm/h][range1][mitigation 1][range 2][mitigation 2][range3][mitigation 3][fog/visibility m][B]

[order id], *Weather Update, 7,[PDI Id], [time], [location], [wind speed: m/s], wind direction, degrees], [rain intensity mm/h], snow intensity [mm/h], [range 1, km], [mitigation factors 1, 0.0-1.0], range 2, km], [mitigation factors 2, 0.0-1.0],[range 3, km] [mitigation factors 3, 0.0-1.0], [visibility, m], [free text] $

Code 38 River Report
Report on River status during the crisis

[report id][type][pdi_id][time][location][water level][water limiti] [B]

[order id], *Weather Update, 2, [PDI Id], [time], [location], [water level, m], [water limit, m], [free text] $

Code 39 People Report
Report on Refugees, Evacuated People, etc.

[report id][source_id][type of report]time][location][number of refugees][adult males][adult females][young people][people with impediments][Number of Red Tags][Number of Yellow Tags][B]

[order id], *Crisis Report, [source_id],[type report: Refugees, Evacuation from Crisis Area Level A, Evacuation from Crisis Area Level B, Evacuation from Crisis Area C], [time], [location], [number of refugees],[adult males],[adult females],[young people],[people with impediments], [Number of Red Tags], [Number of Yellow Tags], [Free Text]$

Code 40 Medical Evacuation Report
Report on Injured People and casualties as well as on arrivals on hospital

[report id][source_id][type of report]time][location][Number of Red Tags][Number of Yellow Tags][Number of Green Tags] [Number of Wounded][Number of Casualties][B]

[order id], *Medevac Report, [source_id], [time], [location], [Number of Red Tags], [Number of Yellow Tags], [number of Green Tags],[Number of Wounded People], [Number of Deads][Free Text]$

Code 41 Single People Report
List of people during the simulation.

[report id][peo_id][time][location][health status][age][gender][impedements][condition][fullname][education level][social status][religion][ethic group][tribe][political party][B]

[order id], * Singel People Report, [peo_id], [time], [location],
[health status: 0-100],[age],[gender: 1 male, 2 female]],[impedements: invalidity percentuage][condition: regular, to be evacuated, escaped],[fullname],[education level: String][social status: String][religion:0-8][ethic group: String][tribe: String][political party: String]
[free text] $

42 Flooded Area Report
Report on flooded areas including contamination level

[report id][type][pdi_id][time][corner1][corner 2][water level][water type] [B]

[order id], *Flooding Update, 2, [PDI Id], [time], [corner1], [corner2] , [water level, m], [water type: clean-mud-contaminated-highly contaminated], [free text] $

CRISOM Input Data
Please check the Annex Directories with file examples

DEM/DTM File
Digital Elevation/Terrain Model.
Format: TIFF (GeoTIFF)
CRS (Coordinate Reference System): EPSG:4258
Step: < 20m

Population File
Urbanistic Unit Population.
Format: csv
Structure:

ID (1)	Name	Sex (2)	Nationality (3)	Age 0-4	Age 5-9	...	Age over 100	Total	Average Age
1	Crevari	M	POLAND	2	2	...	0	8	43,5

(1) Urbanistic Unit ID (only one for each unit)
(2) Value: M / F
(3) Value From the list of Nations (e.g. AFGHANISTAN, ALBANIA, ALGERIA, ANGOLA, …, ARGENTINA, ARMENIA, AUSTRALIA, AUSTRIA, …, BELGIUM,....etc.)

Educational Level File
Number of people for educational Level
Format: csv
Structure:

Country (1)	Country (2)	Year	Population 0-14 years old	-	Primary School	Secondary School	University
ARGENTINA	ARG	2015	10003060	630038	11750805	16010261	4066270

(1) From the Country List
(2) Argentina -> ARG

GDP per capita
GDP per capita for each country
Format: txt
Structure:
ARGENTINA $20,100 (2016 est.)

GINI Index
Gini index for each Country
Format: txt (ASCII)
Structure:
ARGENTINA 42.7 (2014)

Infrastructures Description
Critical Infrastructure
Format: csv
Structure:

Name	Address	Type (1)	Longitude	Latitude	People Capacity	Class (2)	Jobs Capacity
I.C. HYSTORIC CENTER	PIAZZA DI S.MARIA IN VIA LATA, 12	KINDERGARTEN	8.935974	44.40246	70	average	20

(1) The type: "Kindergarten", "Primary School", "Secondary School", "High School", "Private Kindergarten", " Private Primary School ", "Private Secondary School", "Private High School", "University", "museum", "park", "villa", "forte", "theater", "cinema", "beach", "caffe", "stadium", "shop", "trade center", "hospital", "police station", "fire station", "bank", "parking"
(2) "low", "average", "high"

Optional Fields:

Hierarchy	Municipality	Province	Phone Number	Area	District
N.D.	Genova	GE	N.D.	N.D.	N.D.

Natality
Sons Number per woman for each Country
Format: csv
Structure:
ARGENTINA,ARG,2015,2.3080001

Political Preferences
The preferred Political Parties (Chance to vote) for Sex (M/F), age groups (25-29, 29-34…), educational level
Format: csv
Structure:

	Party 1	Party 2	Party 3	Party 4	Party 5	Party 6	Party 7
M	0.1	0.1	0.1	0.1	0.4	0.1	0.1
F	0.4	0.1	0.1	0.1	0.1	0.1	0.1
0-4	0	0	0	0	0	0	0
...							
50-54	0.1	0.2	0.1	0.2	0.1	0.1	0.2
...							
<14 years old	0	0	0	0	0	0	0
-	0.1	0.2	0.1	0.2	0.1	0.1	0.2
Primary School	0.1	0.2	0.1	0.2	0.1	0.1	0.2
High School	0.1	0.2	0.1	0.2	0.1	0.1	0.2
University	0.1	0.2	0.1	0.2	0.1	0.1	0.2

Population Distribution

Population distribution in the urbanistic unit (optional), if the units are not included in the file, or the file miss, the population (houses distribution) is consider uniform.

Format: JSON

Structure:

{"features":[
{"id":29,"name":"Angeli","population_center":[8.906644,44.415041], "population_sigma":[1.0,0.6]},
{"id":58,"name":"Apparizione","population_center":[8.993217,44.403883], "population_sigma":[0.2,0.5]}]}

The Fields:

- **id:** urbanistic unit ID
- **name:** urbanistic unit name
- **population_center:** the coordinates of the unit point with the maximum population density
- **population_sigma**: population distribution efficacious value

Population Incomes

Urbanistic unit average income

Format: csv

Structure:

Urban Unit, Average Income

Example:

1,21.34
2,19.93
3,16.446
4,19.803
5,19.165
6,24.061

Public Events

Week Events affecting Population

Format: csv

Structure:

Type	Kind of Place	Longitude	Latitude	Day of the Week	Starting Hour	Duration [hours]	Description
entertainment	stadium	8.95246	44.41645	6	12:00:00	2	football event

Religion Compatibility

Compatibility defining the chance to be friends or create family

Format: csv

Structure:

	CHRISTIAN	MUSLIM	UNAFFIL	HINDU
CHRISTIAN	1.0	0.5	0.5	0.5
MUSLIM	0.5	1.0	0.5	0.5
UNAFFIL	0.5	0.5	1.0	0.5
HINDU	0.5	0.5	0.5	1.0

Religions

Country Population Percentage of each religion.

Format: txt

Structure:

COUNTRY, PERCENT CHRISTIAN, PERCENT MUSLIM, PERCENT UNAFFIL., PERCENT HINDU, PERCENT BUDDHIST, PERCENT FOLK RELIGION, PERCENT OTHER RELIGION, PERCENT JEWISH

Example

AFGHANISTAN,0.1,99.7,0.1,0.1,0.1,0.1,0.1,0.1
ALBANIA,18,80.3,1.4,0.1,0.1,0.1,0.2,0.1
ALGERIA,0.2,97.9,1.8,0.1,0.1,0.1,0.1,0.1
….

Rivers & Sewerage

Rivers, Aquifer, Sewerage and Manholes of the scenario

Format: JSON

Structure:

{ "node_id": 0,
"node_coordinates": [
44.39431220,
8.942613601
], "node_links": [
{
"link_id": 7
}] }, {
"node_id": 1,
"node_coordinates": [
44.407359497,
8.9502578973
],
"node_links": [
{
"link_id": 7
},
{
"link_id": 8
},
{
"link_id": 114
}] }],
"links": [
{ "link_id": 0,
"node_1": 18,
"node_2": 20,
"depth": 4,
"width": 50,
"type": "normal",
"distance": 303.501
},
{
"link_id": 1,
"node_1": 180,
"node_2": 79,
"depth": 1,
"width": 3,
"type": "underground",
"distance": 1161.451
}] }

The Fields:

node_id: node ID
node_coordinates: node coordinates
node_links: links connected to the node
link_id: link ID
node_1, node_2: the nodes connected by the link
depth: average depth of the river

width: average width of the river bed
type: link type (normal/underground)
distance: distance of the link in meters

Roads

The roads of the scenario
Format: JSON
Structure:

```
{  "nodes": [
   {      "node_id": 0,
     "node_coordinates": [
       44.41520655,
       8.707437515
     ],
     "isolated": false,
     "node_links": [
       {
        "link_id": 195
       },
       {          "link_id": 225
       }       ]    },
   {      "node_id": 1,
     "node_coordinates": [
       44.42575053,
       8.736963272      ],
     "isolated": false,
     "node_links": [
       {
        "link_id": 10
       },
       {
        "link_id": 195
       }       ]     } ],
"links": [
   {      "link_id": 281,
     "node_1": 55,
     "node_2": 62,
     "type": "normal",
     "distance": 3525.494,
     "direction": 0
   },
   {
     "link_id": 1,
     "node_1": 200,
     "node_2": 88,
     "type": "normal",
     "distance": 182.210,
     "direction": 0
   }  ]  }
```

The Fields:

node_id: node ID
node_coordinates: node coordinates
node_links: vector of links connected to the node
link_id: link ID
node_1, node_2: the nodes connected by the link
type: type of the road (normal/highway/railway)
direction: direction of the traffic:
0 - bidirectional
1 - from node 1 to node 2
-1 - from node 2 to node 1

Terrain Type

Terrain Type
Format: TIFF (GeoTIFF)
CRS (Coordinate Reference System): EPSG:3035
The Terrain Type must be compatible with the format used in the Copernicus Project (www.copernicus.eu)

Urban Areas

Urbanistic unit (polygons)
Format: GeoJSON
Structure:

```
{"type":"FeatureCollection","features":[{"type":"Feature","geometry":{"type":"Polygon","coordinates":[[[8.83628,44.423768,0],[8.83644,44.423748,0],[8.836494,44.423726,0],[8.836726,44.423661,0]]]},"properties":{"stroke":"#555555","stroke-width":2,"stroke-opacity":1,"fill":"#555555","fill-opacity":0.5,"name":"Porto","id":0}}]}
```

Weather Forecasts

Weather Forecast
Format: csv
Structure:

Weather Station	Date & time	Rain intensity [mm/h]	Temperature [C]	Wind speed [m/s]	Wind Direction
1	12 Feb 2018 07:00:00	10	12	3	90

Weather Statistics

Typical values of the temperature, wind direction
Format: csv
Structure:

Weather Station (1)	Temperature Correction Hourly (2)	Wind Direction (3)	Wind Speed (4) [m/s]
1, Christopher Columbus, Airport, 8.847598, 44.414878	-3.6, -3.6, -3.6, -3.4, -3.7, -3.7, -2.8, -1.7, -0.6, +1.0, +3.2, +3.4, +3.7, +2.9,+1.8, +1.7, +1.6, +0.6, -0.6, -1.4, -2.7, -3.2, -3.5, -3.7	270, 180, 090, 090, 090, 090, 090, 090, 090, 180, 180, 270	6, 6, 5, 4, 3, 2, 1, 1, 4, 5, 5, 5

1) Weather Station that gives the weather forecast:
ID, name, description, longitude, latitude
2) Change in temperature during a day compared to average temperature
3) Typical Wind Direction in each month of the year
3) Typical Wind Speed in each month of the year

Work and Commercial Activities

Workplace and attractiveness number of each urbanistic unit
Format: csv
Structure:

Unit id	Unit Name	Workplace High	Workplace Average	Workplace Low	Attraction High	Attraction Average	Attraction Low
0	Porto	750	1800	450	1.5	2	2

The Fields:

Id and *Name* of urbanistic unit
Workplace number of employed in this area subdivided in terms of income groups: low/average/high
Attractiveness of urbanistic unit: 0 – doesn't attract, 1 – average level, >1 - high level

CONCLUSIONS

The paper proposes an effective use of simulation to address crisis impacting urban areas in relation to critical infrastructure and industrial facilities; it is proposed an innovative agent-driven interoperable hybrid stochastic simulation and data structures to support interoperability and connection with other systems and simulators. The proposed approach represents a good opportunity to use simulator in synergy with other models, tools and systems to enhance training, planning, operational support as well as safety engineering. Currently the authors are using it for multiple case studies, extending further its potential and conducting additional experimentation on interoperability with other systems.

REFERENCES

Acquaotta, F., Faccini, F., Fratianni, S., Paliaga, G., & Sacchini, A. (2017). Rainfall intensity in the Genoa Metropolitan Area (Northern Mediterranean): secular variations and consequences. Weather.

Bruzzone A.G., Maglione G.L., Agresta M., Di Matteo R., Nikolov O., Pietzschmann H., David W. (2017a) "Population Modelling for Simulation of Disasters & Emergencies Impacting on Critical Infrastructures", Proc.of WAMS, Florence, September

Bruzzone A.G., David W., Agresta M., Iana F. Martinesi P., Richetti R. (2017b) “Integrating Spatial Analysis, Disaster Modeling and Simulation for Risk Management and Community Relisience on Urbanized Coastal Areas”, Proc. of 5th Annual Interagency, Interaction in Crisis Managements and Disaster Response, CMDR, June 1-2

Bruzzone, A.G., Agresta, M., Sinelshchikov, K. (2017c). Simulation as decision support system for disaster prevention. SESDE 2017, Barcelona, Spain

Bruzzone A.G., Longo F., Massei M., Nicoletti L., Agresta M., Di Matteo R., Maglione G.L., Murino G., Antonio Padovano A. (2016a) “Disasters and Emergency Management in Chemical and Industrial Plants: Drones simulation for education & training”, Proc. of MESAS, Rome, June 15-16

Bruzzone A.G., Massei M., Di Matteo R., Agresta M., Franzinetti G., Porro P. (2016b) “Modeling, Interoperable Simulation & Serious Games as an Innovative Approach for Disaster Relief”, Proc.of I3M, Larnaca, Sept. 26-28

Bruzzone A.G., Cirillo L., Longo F., Nicoletti L.(2015) “Multi-disciplinary approach to disasters management in industrial plants”, Proc.of the Conference on Summer Computer Simulation, Chicago, IL, USA, July, pp. 1-9

Bruzzone A.G., Frascio M., Longo F., Chiurco F., Zanoni S., Zavanella L., Fadda P., Fancello G., Falcone D., De Felice F., Petrillo A., Carotenuto P. (2014) “Disaster and Emergency Management Simulation in Industrial Plants”, Proceedings of EMSS2014, Bordeaux, France, September

Bruzzone, A. G. (2013). Intelligent agent-based simulation for supporting operational planning in country reconstruction. International Journal of Simulation and Process Modelling, 8(2-3), 145-159.

Bruzzone A.G., Scavotti, A., Massei, M., Tremori, A.(2008) "Metamodelling for analyzing scenarios of urban crisis and area stabilization by applying intelligent agents", Proceedings of 20th European Modeling and Simulation Symposium, Amantea, Italy, September 17-19, pp.521-526

Bruzzone A.G., Giribone P., Mosca R. (1996) "Simulation of Hazardous Material Fallout for Emergency Management During Accidents", Simulation, vol. 66, no.6, June, 343-355

Chanson, H. (2004). Hydraulics of open channel flow. Elsevier.

Duley, F. L., & Kelly, L. L. (1939). Effect of soil type, slope, and surface conditions on intake of water.

Faccini, F., Luino, F., Paliaga, G., Sacchini, A., Turconi, L., & de Jong, C. (2018). Role of rainfall intensity and urban sprawl in the 2014 flash flood in Genoa City, Bisagno Catchment (Liguria, Italy). Applied Geography, 98, 224-241.

Harding J. (2018)“CCTV shows Genoa bridge collapse”, BBC News, August 21

Kuhl, F., Dahmann, J., & Weatherly, R. (2000). Creating computer simulation systems: an introduction to the high level architecture. Upper Saddle River: Prentice Hall PTR.

Longo, F. 2013. On the short period production planning in industrial plants: A real case study. International Journal of Simulation and Process Modelling, vol. 8, no. 1, pp. 17-28.

Mengfan Z. (2016) “Tianjin chemical blast: Deadly disaster marked one year on”, BBC News, August 12.

Ören, T., Longo, F., 2008. Emergence, anticipation and multisimulation: Bases for conflict simulation. 20th European Modeling and Simulation Symposium, EMSS 2008, pp. 546.

<u>Author's index</u>

ISBN 978-88-85741-12-6; Bruzzone and Sottilare Eds.