# Modeling and Simulation of Red Teaming[1]

## Part 1: Why Red Team M&S?

Michael J. Skroch, Sandia National Laboratories

2 November 2009 – Rev 3

### Introduction

Red teams that address complex systems have rarely taken advantage of Modeling and Simulation (M&S) in a way that reproduces most or all of a red-blue team exchange within a computer. Chess programs, starting with IBM's Deep Blue, outperform humans in that red-blue interaction, so why shouldn't we think computers can outperform traditional red teams now or in the future? This and future position papers will explore possible ways to use M&S to augment or replace traditional red teams in some situations, the features Red Team M&S should possess, how one might connect live and simulated red teams, and existing tools in this domain.

### Why Red Team M&S?

Successful Red Teaming is often all about the individual or team that has been pulled together to perform the job at hand. Mission Impossible's Mr. Phelps regularly went through the dossiers of potential team members to find just the right ones to do the job. Similarly, many successful efforts in actual sophisticated red teams have a lot to do with team composition. Yet, this is just one of the tools that Mr. Phelps had at hand. Others included specific technologies, the right information, and ability to socially engineer the target humans in the objective.

Mr. Phelps had a particular luxury that many users of red teams and red teams themselves don't have – he wasn't leading a red team, he was leading an aggressor squad with a focused target and objective. Real red teams are an extension of a defensive strategy for organizations, companies, and governments that must consider a breadth of attacks from their own postulated aggressors—they have to consider a broad set of possible attacks, not just one. Methods used by that red team may span physical attacks, cyber attacks,

#### Definitions* & Focus

Many terms have particular meaning for their community. For this article series, here's how I will focus:

*Red teaming* is an adversarial-based assessment of your security, plans, strategy or other system that may be prone to a malevolent threat. In the context of this article, it is focused on a complex system that includes two or more attributes of physical, cyber, and human behavior.

Simulation, in this article, refers to a Modeling & Simulation (M&S) approach that is computer-based in some aspect but may include Live, Virtual, and Constructive elements (LVC). Most often we are referring to constructive simulations that do not include real or live humans and technology.

A model, in this article, implies a computer model or mathematical or data-driven representation of an object or process.

Red Team M&S is the practice of reproducing most or all of a red-blue engagement with a computer simulation, particularly on systems for this paper.

M&S for Red Teaming is the practice of a red team to utilize M&S as a tool within its overall live red team activity from planning to execution.

* Definitions here are compiled from a wide variety of sources and author experience. They are intended primarily to provide focus to the position discussion.

and manipulations of human staff. So, how is a red team or set of red teams supposed to cover the broader ground of a defensive entity and emulate all likely attacks on their target system? Providing more time and more funding are not popular answers to this question. The only real answer is effectiveness directed by a cost-benefit approach to how the red team will cover possible aggressor actions.

Why focus upon effectiveness? Because use of adversarial perspective in design is essential and the ground red teams must cover is growing. System security is not keeping pace with threats. Of equal importance is the impracticality of staffing and funding a sufficient number of red teams to address the problems we now face or we project into the future.

Effectiveness for red teams may come in many forms. A defined process or approach can improve accuracy and completeness of a red team. One that can be trained to a broad set of individuals will allow a broader army of red teamers following that process. Technologies enable red teams with capabilities that didn't exist or were previously impractical. Some process and technology enables those less initiated to emulate those with more skill and experience. One relatively new and very promising tool for red teaming effectiveness is modeling and simulation.

## M&S cannot do everything. What can it do?

In researching this topic, I've discussed modeling and simulation OF red teams with red team members of various types. I often receive an immediate, almost guttural, push-back that M&S cannot replace a live red team. Nobody wants to become obsolete because they were replaced by a machine or computer program. Yet, in this case their assertions seem plausible. A live red team has features that cannot currently be matched or replaced by a computer simulation. I would never advocate that a simulation can replace a red team, particularly to a well-armed Special Forces red team member that has no problem with direct expression. With that exception aside, this brings up a number of interesting questions for discussion: *When and in what ways is a simulation more useful than a live red team? How can simulations of red teaming be used in concert with live red teams? How do you reliably decide which to choose? Are there situations where M&S can replace a red team?*

One could probably answer quickly that humans are good at doing what humans can do (intuition, creativity, etc.) and computers are good at doing what computers can do (complexity, crunch numbers). Where else have computers done well and are mature in simulating red-blue interactions? Consider a few:

- Chess
- Video games
- Economic models
- Ecological models

Common features of all these include a well-defined environment in which the programmers and mathematicians have had a long time to model the environment or analyze the system. These examples also consider systems that are innately simple or can be reduced in complexity for simulation. Complex red team situations involving physical, cyber, and behavior do not share these features.

## Close but not systems Red Teaming

Both M&S and Red Teaming are broad domains. In order to narrow the focus of this paper, I want to acknowledge but set aside fields that are topically near or overlap with the focus of this paper.

A growing body of work involves research and development of adversarial behavior modeling. A number of these involve simulation for training. A good Red Team M&S system should benefit from this research and be a source of adversarial stimulus for such training; however, I wish to focus on red teaming of systems. A potential way to distinguish training simulation from systems Red Teaming M&S

is that the latter can operate without human participation, that is, operate in a constructive mode.

There is a centuries-long military history of force-on-force modeling and non-computer simulation (often called wargaming) for tactics development and training that has lead to the inevitable inclusion of computers for this pursuit. Most of these efforts focused upon training in flight or on the battlefield[2]. Some recent research and development level training systems that exist include DARWARS Ambush! Trainer[3], the RealWorld[4] system from DARPA, and Ground Truth[5] from Sandia. We will exclude such systems from consideration in this paper.

Another set of modeling and simulation tools that I will exclude for this paper is that which analyzes scenarios, but without some level of adaptive human behavior. They tend to evaluate system-on-system performance without respect to complex human interplay. Some may model human decisions, attack graphs, or defensive tactics without actually performing any force-on-force interaction. All these tools[6] are more suited to M&S for Red Teaming applications.

Yet another set of simulations I will exclude is that that which models interactions of thousands to millions of agents making national or globally relevant decisions. Interplay is often economic-based and shows trends or effects from impacts to a system of systems. An example of such a capability that often considers consequence-based analysis is the NISAC[7]. While certainly a potential M&S tool for Red Teaming, these types of capabilities do not focus on the embodied interaction of a typical red team force upon a target.

## Defining Red Teaming M&S

With these modeling and simulation topics set aside for now, a high-level positive definition of Red Teaming M&S will be useful for comparing available systems and refining detailed requirements.

Required
- Simulates force-on-force interplay
- Complex adaptive human behaviors
- Simulates 3D physics, cyber, or both
- Includes constructive-only mode

Optional
- Includes virtual and live interplay
- Federates with other simulations
- Embodied human/object behaviors

This working definition implies that the Red Team M&S will be able to constructively simulate human interaction with physical systems, cyber systems or both. Interplay of physical-physical (e.g., explosion, bullet) and physical-cyber (e.g., control systems, physical destruct of information) would also be required in the simulation.

## Red Team Measures of Effectiveness

How are we to provide some objectivity to the use of Red Team M&S? Developing Measures of Effectiveness (MOE) or a framework for comparison is a good start. Consider that a number of simulations may be able to provide capability for force-on-force simulation including OneSAF[8], JSAF[9], STAGE[10], Simajin[11], Avert[12], Umbra[13], Dante[14]. Other tools may also be useful for this purpose. I'll explore some of those in a future paper in this series—first

[2] For example, DARPA's SIMNET and the Army's Close Combat Team Trainer (CCTT)..
[3] www.darwars.net, en.wikipedia.org/wiki/DARWARS/; Ambush trainer from SNL, Elaine Raybourn.
[4] DARPA DSO Office, www.totimm.com
[5] Ground Truth incident responder training game, Sandia National Laboratories, Donna Djordjevich.
[6] Examples include fault tree tools, path planning tools like ASSESS by Sandia.
[7] National Infrastructure Simulation and Analysis Center (NISCA), Department of Homeland Security (DHS).
[8] One Semi-Automated Forces (OneSAF), www.peostri.army.mil/PRODUCTS/ONESAF/
[9] Joint Semi-Automated Forces (JSAF), predecessor of OneSAF, www.jfcom.mil/about/fact_jsaf.html.
[10] STAGE, AI.implant, etc., Presagis Inc., presagis.com.
[11] Simajin, RhinoCorps LTD, rhinocorps.com.
[12] Automated Vulnerability Evaluation for Risks of Terrorism (AVERT), ARES Corporation, www.arescorporation.com.
[13] Umbra Simulation Framework, Sandia National Laboratories, umbra.sandia.gov.
[14] Dante scenario analysis simulation tool, Sandia National Laboratories, umbra.sandia.gov.

we should have a foundation for comparison.

Measures and metrics for red teaming might be split into four categories shown here:

- Target
  - Behavior, consequence, risk, …
- Adversary
  - Behavior, resources, …
- Red team process
  - Development, attack, success, …
- Red team effectiveness
  - Capability, performance, …

Sandia National Laboratories refers to the first three in its processes and courses "Red Teaming for Program Managers" (RT4PM)[15] and "Red Team Metrics."[16] While "adversary" metrics are associated with red team characterization, including progression of effort, they are not focused upon the performance of the red team itself. Qualifying a red team is something the RT4PM process hints at through a-priori measures of effectiveness including experience, composition, process, capability, and knowledge. Beyond this, nothing in these efforts documents a formal consideration for measuring effectiveness of a particular red team or red team simulation before and during a red teaming effort.

University of Wisconsin-Madison conducted a 2004 study[17] of Sandia National Laboratories' Information Design Assurance Red Team (IDART) that was focused upon red team performance and collected various measures of effectiveness. It also devotes one section to "weaknesses and strengths of simulation methods."

A summary of measures of effectiveness is listed here. Notice that these measures focus primarily on the human-centric red team and don't translate well to simulated red teams.

- Team design, composition
- Team member quality
- Material resources
- Team synergy, shared vision, conflict, trust
- Process

A simulated red team will consider the above items as variables in their simulation design while live red teams struggle with these as resources or maturity issues.

Comparisons of simulations to human-based red teams in this work discussed the following characteristics:

- Testing against known issues
- Anticipatory
- Adaptability
- Exhaust range of possibilities
- Environment complexity
- Creativity
- Understanding target goals in larger context
- Ability to analyze, find patterns

This author observed the UW-Madison study noted here and noted that simulations being considered were primarily M&S tools for Red Teaming, not Red Team M&S. Comments about Red Team M&S were speculative in that no holistic computer simulation existed at the time to simulate red team activity, particularly for red teaming information systems. Comments collected in the study show that the red team members believed it was unlikely that a simulation could duplicate a red team, but that simulation offered promise as a tool for red teamers and also had the potential to do what computers do well—crunch numbers and exhaust the range of alternatives.

Given the UW-Madison paper and other experiences with simulation environments, I postulate measures in Table 1 that are important in comparing live red teams with Red Team M&S. All of the UW-Madison measures are included with the exception of the two underlined items. These two deal with systems analysis, which is often part of red team preparation or after-action studies. At this point the measures have no detailed range of value, expected distribution, or weighting. I will attempt to add that detail in a future paper.

[15] http://idart.sandia.gov/training/RT4PM.html
[16] http://idart.sandia.gov/training/Metrics.html
[17] "Red Team Performance: Summary of Findings; University of Wisconsin-Madison & IDART: Sandia National Laboratories," June 2004, Pascale Carayon and Sara Kraemer, University of Wisconsin Center for Quality and Productivity Improvement.

| Measures | Live Red Team | Red Team M&S |
|---|---|---|
| Adaptation, agility, unknowns, creativity | Inherently adaptable within skill set and resources of team. | Immature domain for simulation. Adaptable within limitations of programming. |
| Breadth of knowledge | Inherently broad range within team or resources available to team. | Limited to that which is provided to and can be effectively used by the simulation. |
| Fidelity, precision | Inherently broad within skill set and tools available. | Adjustable based upon capability of simulation and ability to model. Potential for fidelity exceeding human abilities. |
| Learning and unlearning | Inherent ability to learn. Difficult to "unlearn" or forget, thus causing tainting of future red team efforts. | Ability to learn based upon fidelity of model. Ability to "unlearn" or forget what has been done in the past. |
| Stochastic variation, range of possibilities | Variation can be expensive due to labor costs and limits on time to reproduce the red team effort. | Implicitly possible to create a wide range of variation quickly and at low cost. |
| Live Virtual Constructive (LVC) | Primarily focused on live. Ability to use tools to engage cyber systems. Inability to truly engage LVC without simulation. | Some systems allow LVC, primarily focus on LV, with few incorporating C. Merging live and constructive red team simulation has potential. |
| Measureable results, data collection | Often difficult depending on how the red team is instrumented. May slow the red team effort or increase costs. | Implicitly available, all data is usually available and can be recorded with little cost. |
| Speed, capacity | Usually limited to real time, with exception based upon speculative tools for red team. Limited by number of read team. | Essentially unlimited. Essentially limited only by computing resources available. |
| Reproducible results | Requires use of methodology and proper selection of team members. | Inherently able to reproduce past events and provide consistent environments for constructive simulations. |
| Accurate results – ability to be validated and verified, V&V | Requires rigor of process, shadow red team, multiple red teaming, V&V must be reconsidered with each new team | Nature of simulation enables V&V that is sustained across similar simulations or models |
| Cost | Usually considered higher cost due to labor. May cost less in a small tactical red team effort. Depends on breadth and depth. | Potential to reduce costs for complete coverage of attack spaces, stochastic variations, sensitivity analysis, etc. May have higher cost if only used for a single run. |
| Time to set up | Depends on size of red team assembled, time to assemble resources for exercise. Scales linearly. | Depends on M&S tool features, architecture, detail of simulation required. Amortizes over number of runs. |
| System breadth, complexity, entities | Limited by team size, ability to keep data in mind, time and ability to collect data. | Limited by particular M&S system constraints, time and ability to collect data. |
| Ability to federate | Innate within constraints of communication. | Dependent on particular M&S system. [HLA, DIS, TENA...] |
| Bias, COI[18] | Depends on individuals, team affiliation. | Depends on affiliation of analysts, programmers. |
| Efficiency as a cost-benefit value | Based upon quality of process, team composition. | Potential to be more highly efficient than a red team due to automation. |
| Effectiveness (w/o regard to efficiency) | Depends on domain of use. | Depends on domain of use. |
| Physical RT [gates, guns, guards, …] | Mature discipline requires knowledgeable team members, can be highly effective but not necessarily exhaustive. | Physics-based nature of physical systems lends well to M&S. Ability for excellent variable fidelity results and exhaustive results. |
| Cyber RT [hardware, software, network, …] | Discipline immature and often driven by team member expertise, access to particular tools, resources | Isolated hacking and fuzzing tools exist with increasing "intelligence" but are primarily scripted. Potential for speed. |
| Behavioral RT [skill, culture, M&I[19], psyop, phishing, …] | Live humans are the benchmark for behavior in live systems. | Maturing discipline, behavioral modeling is in its infancy. Able to model defined tactics, techniques and procedures with some success. |

Table 1: Postulated Measures of Effectiveness for Red Teams and Red Team M&S

[18] Conflict Of Interest (COI)
[19] Motivation & Intent (M&I)

For each measure in Table 1, I provide comment for live red teams and simulation of red teaming based upon my experience and discussions with others. I've expressed my opinion on which category currently "wins" with respect to effectiveness in red teaming. Those highlighted green are the "winners." Those rows with no highlight seem to be tied or too close to tell depending on the red team situation. This speculation is based upon experience with hundreds of red team activities performed by IDART and other red team accounts across many domains and customers.

Another discussion on this topic comes from a conversation on Red Team Journal, "What Factors Characterize Successful Red Teaming?"[20] that took place in July-August, 2009. Factors seem to distill down to team composition, credibility (trust), process (many), knowledge of target, and independence (lack of bias, lack of conflict of interest).

Upon first glance of Table 1, my bias toward the need for Red Team M&S may seem apparent because more columns win for simulation. Note that it is not my intent to imply that simulation is always a best approach. Here are a few reasons why:

1. This is a high-level specification that may not match specific red team interests or scenarios.
2. Each row does not provide the same weight of evidence for effective red teaming. For instance, the first two rows, adaptation and breadth of knowledge, may be the principle values desired in red teaming and easily outweigh the others.
3. Columns that focus upon simulation represent technology or approaches that have traditionally been overlooked or not applied by live red teams. Therefore existing capability in these areas is currently low for live red teams and would bias the answer toward M&S.

[20] http://redteamjournal.com/2009/07/what-factors-characterize-successful-red-teaming/

## Physical – Cyber – Behavior

The last three columns of Table 1 list ability of the red team to work in the three domains of physical (3D environment, physical weapons), cyber (computers, networks, information), and behavior (human, societal, cultural, policy). These three domains were chosen as a means to cover system problem space because they are fairly well-known and map well to science and engineering disciplines. Ability in each domain is important, but ability to integrate two or more domains is crucial in order to address concepts of systems-of-systems and interdependence.

This relates directly to comments in the UW-Madison study that express need for "Environment complexity." Live red teams obviously have innate ability to perform this integration, but will be limited based upon the team and its resources. Simulation will have advantage and disadvantage in ways previously discussed.

## Feasibility for Red Team MOEs

To appease critics of security metrics development, I should point out that many articles[21] discuss difficulty of applying metrics to security, so I mention here that viability of metrics in Table 1 is not guaranteed. Furthermore, I point out that this particular quest for metrics is not fully intertwined with the security metrics debate. There is hope given that we're measuring a tangible red team or M&S application rather than the broad concept of a singular system security metric.

[21] For instance, "Security Metrics for Communication Systems," Mark D. Torgerson, Sandia Natinoal Laboratories, 2008, 12th ICCTRS –and– "Security Metrology and the Monty Hall Problem," Bennet S. Yee, April 2, 2001.

## Interim conclusions, next steps

With this foundation of concepts and speculation defined, I can summarize my position on this topic:

- Red Team M&S does not replace red teams; it augments that practice by providing tools to both analysts and red teams.
- Red Team M&S has the potential to capture some red team knowledge and apply it more broadly and at less expense than live red teams.
- Red Team M&S can cover a wider set of possibilities, thus directing a live red team where it is needed.
- Red Team M&S naturally extends the measures of performance that are required for considering how to use red teaming or particular red teaming tools.

Additional detail is needed. Measures provided in Table 1 are generalized and need more detail with respect to various types of red teaming and various types of M&S. There are also a number of questions remaining including these:

- When should you use a red team and Red Team M&S together or separately?
- Is a live red team required to set up Red Team M&S?
- What is the applicability of Red Team M&S to various points in a system's lifecycle?
- What portions of a live red team lifecycle or set of activities can Red Team M&S duplicate?
- Can an analyst plus a Red Team M&S tool duplicate much of a live red team effort?
- What features should Red Team M&S have for various types of systems and red teaming?
- Are there benefits to having live red team operations augmented by real-time M&S?

Future position papers on topic this will address these and other thoughts about Red Team M&S. Comments about the topics in this paper are welcome at mjskroc@sandia.gov. Videos of some Umbra/Dante capability and efforts in this area can be seen on youtube.com by searching for "umbrasandia."

## Author Biography

Michael J. Skroch (skraw) has been at Sandia National Laboratories for 22 years and is currently a Manager of the Interactive Systems Simulation & Analysis Department, which focuses on modeling and simulation as a tool for analysis, assessment, and training in combined physical, behavior, and cyber environments. He established leading red teams in the nation at Sandia including the Information Operations Red Teaming and Assessments (IORTA) area and Information Design Assurance Red Team (IDART). He has served at DARPA as a Program Manager and the Pentagon in the area of Information Assurance.