

Complex Systems and Human Performance Modeling

by Walter Warwick, Laura Marusich, and Norbou Buchler

ARL-RP-467

December 2013

*A reprint from Proceedings of the Human Factors & Ergonomics Society 2013 Annual Meeting,
San Diego, CA, 30 September–4 October 2013.*

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5425

ARL-RP-467**December 2013**

Complex Systems and Human Performance Modeling

Walter Warwick

Alion Science and Technology

Laura Marusich and Norbou Buchler

Human Research and Engineering Directorate, ARL

*A reprint from Proceedings of the Human Factors & Ergonomics Society 2013 Annual Meeting,
San Diego, CA, 30 September–4 October 2013.*

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) December 2013		2. REPORT TYPE Reprint		3. DATES COVERED (From - To) January 2012–November 2013	
4. TITLE AND SUBTITLE Complex Systems and Human Performance Modeling				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Walter Warwick,* Laura Marusich, and Norbou Buchler				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory ATTN: RDRL-HRS-E Aberdeen Proving Ground, MD 21005-5425				8. PERFORMING ORGANIZATION REPORT NUMBER ARL-RP-467	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES * Alion Science and Technology, 4949 Pearl East Circle, Ste. 200, Boulder, CO 80301 A reprint from <i>Proceedings of the Human Factors & Ergonomics Society 2013 Annual Meeting</i> , San Diego, CA, 30 September–4 October 2013.					
14. ABSTRACT The development of a human performance model is an exercise in complexity. Despite this, techniques that are commonplace in the study of complex dynamical systems have yet to find their way into human performance modeler's toolbox. In this report, we describe our efforts to develop new generative and analytical methods within a task network modeling environment. Specifically, we present task network modeling techniques for generating inter-event times series typical of a complex system. We focus on communication patterns. In addition, we describe the associated analytical techniques needed to verify the time series. Again, while these analytical techniques will be familiar to the complexity scientist, they have significant and largely unrecognized methodological implications for the human performance modeler.					
15. SUBJECT TERMS human performance modeling, network science, C3TRACE, power laws					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 12	19a. NAME OF RESPONSIBLE PERSON Norbou Buchler
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include area code) 410-278-9403

COMPLEX SYSTEMS AND HUMAN PERFORMANCE MODELING

Walter Warwick
Alion Science and Technology

Laura Marusich & Norbou Buchler
Human Research & Engineering Directorate, US Army Research Laboratory

The development of a human performance model is an exercise in complexity. Despite this, techniques that are commonplace in the study of complex dynamical systems have yet to find their way into human performance modeler's toolbox. In this paper, we describe our efforts to develop new generative and analytical methods within a task network modeling environment. Specifically, we present task network modeling techniques for generating inter-event times series typical of a complex system. We focus on communication patterns. In addition, we describe the associated analytical techniques needed to verify the time series. Again, while these analytical techniques will be familiar to the complexity scientist, they have significant and largely unrecognized methodological implications for the human performance modeler.

INTRODUCTION

The development of a human performance model is an exercise in complexity. Whether it is instantiating the various mechanisms that constitute a cognitive architecture or decomposing the work flows and resource constraints that characterize human-system interactions, the modeler confronts the challenge of representing and understanding the myriad interactions that drive human performance in a dynamic environment. Even the modeling tools themselves are instances of complex software systems. Despite this, techniques that are commonplace in the study of complex dynamical systems have yet to find their way into the human performance modeler's toolbox.

In this paper we describe our efforts to implement methods for generating and analyzing the inter-event times series typical of a complex system within a human performance modeling environment. More specifically, we describe how the stochastic priority-based queuing proposed by Barabasi (2005) to account for the heavy-tailed distribution of wait times in human communication patterns can be implemented in a task network modeling tool. Although queues are a basic feature in many task network modeling tools, the representation of Barabasi's stochastic priority-based queuing is rather more subtle than it might first appear. Nor is it enough simply to represent the functional form of the heavy-tailed distribution explicitly in the calculation of task times. Rather, our goal to demonstrate how a heavy-tailed distribution of wait times can *emerge* having re-represented stochastic priority-based queuing as a form of task switching. In this way we reconcile what some might see as the more recondite concepts of the complexity scientist with the more familiar terms of the human performance modeler.

Given our goal to generate an emergent time series, it is necessary to implement analytical techniques to identify the functional form of the resulting distribution of wait times. Again, this problem is more subtle than it appears. Clauset, Shalizi and Newman (2009) point out that the common approach of simply displaying linear fits to log-log plots of the distribution is fraught with problems. So, following their lead,

we have implemented general techniques for verifying power-law distributions in data (empirical or simulated).

Our ultimate goal in this paper is not to argue whether a particular aspect of human performance is best described by, say, a power-law or a stretched exponential or a log-normal distribution (cf. Anderson, 1981; Heathcote, Brown & Mewhort, 2000). Rather, we think it is important to understand the methodological implications should an aspect of human behavior be driven by a power law process. Our most immediate concern has to do with the role of computational modeling in predicting system performance under such a regime. One of the key findings emerging from network science is that the statistics of complex systems are "non-normal" (West & Grigolini, 2011) while most human performance modeling tools assume some degree of "normality" in the processes they model, if not by design then certainly in their typical application. A model built on draws from normal distributions will fail to account for the extreme behaviors of a system that is, in fact, driven by a power-law process. A more fundamental concern relates to how we should validate models of human performance. While some might view questions about the exact functional form of a distribution as an esoteric exercise in curve fitting, the statistics of complex systems preclude standard measures of fit (and it is for this reason that arguments about the exact functional form of a distribution must be engaged carefully lest they become question begging). Last, but not least, we hope this research will foster stronger connections between human performance modeling and network science—two disciplines that should be tightly linked in principle but that are rarely integrated in practice.

TRADITIONAL TASK NETWORK MODELING

Task network modeling environments (e.g., IMPRINT, Micro Saint, C3TRACE, etc.) provide a framework for representing complex human-system interaction. The framework is grounded in the notion of a task decomposition in which a complex process is decomposed into component sub-processes, which can be further decomposed until a sequence of tasks can be specified. This approach is visualized

by way of a graphical representation of the process being modeled: a hierarchical set of nodes and edges (i.e., tasks and the flow of control between them). Beneath the graphical representation, a discrete event simulator schedules events and executes whatever additional code the modeler has included at the task level (e.g., code to affect state variables). Both serial and parallel processes can be represented directly (see Figure 5 below).

The task is the atomic unit of analysis and, as such, has some basic properties specified by the modeler. Task duration is chief among these properties. By specifying a duration distribution for each task, the modeler provides the stochastic outline for the time series of events that will be simulated at run time.

MODELING COMMUNICATIVE BEHAVIOR

Barabasi (2010) argues that human communication patterns are “bursty”; that is, the inter-event arrival time between messages tends to follow a power-law distribution, with short intervals between many messages but potentially yawning gaps between others. As Figure 1 shows, there is a marked difference between the range of inter-event arrival times given a power-law process as opposed to a Poisson process.

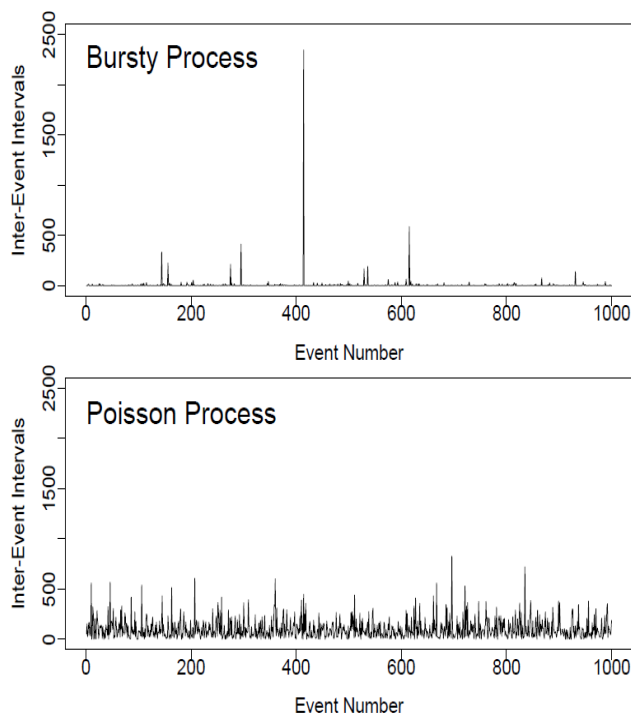


Figure 1 Inter-event arrival intervals given a process governed by a power law (i.e., a “bursty” process) versus inter-event arrivals given a Poisson process. Barabasi (2005) provides a more detailed depiction of this difference.

Barabasi claims that such patterns can be found in both hand-written correspondence (Oliveira & Barabasi, 2005) and email communications (Barabasi, 2005). We have found similar patterns in other data sets. The rank-frequency plot in Figure 2 depicts the distribution of inter-event arrival intervals

for roughly 2300 chat communications from 64 officers arrayed across 15 chat rooms engaged in a two-day command and control exercise. (The data and the exercise from which they were collected are described in Foltz, Lavoie, Oberbreckling, & Rosenstein 2008).

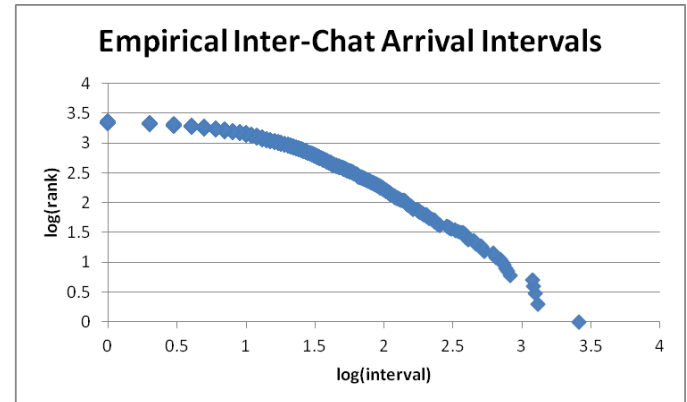


Figure 2 Rank-Frequency plot of the inter-chat arrival intervals from the live C2 exercise.

The linear plot on the log-log graph suggests a power-law distribution and using techniques we describe later, we have estimated a power-law exponent of ~ 2.44 . Although this exponent differs from the value Barabasi reports (~ 1 for email exchanges) the overall communication pattern is qualitatively similar to those he describes. Rather than explain away the differences between email and chat communications we will instead focus on how we might recapitulate the general behavior in a task network model. In particular, we are interested in C3TRACE (cf. Plott, 2003), a task network modeling tool that was developed specifically to simulate the flow of information across command and control structures. The flow of information is represented as the exchange of messages of various types (e.g., e-mail, radio, written etc.) among the personnel in the command structure. The arrival of each message initiates the performance of various tasks which, in turn, drive predictions about performance along a variety of dimensions (e.g., time to completion, workload over time, message throughput etc).

Clearly, the validity of those predictions depends in large part on correctly simulating the arrival intervals between messages. It is less clear is how to generate the underlying time series given the many degrees of freedom available to the C3TRACE modeler. At the most basic level, the modeler must decide whether the message traffic should be fixed as an input to the model (i.e., as a known sequence of events specified prior to run time) or whether the messages traffic will be generated at run time in response to other dynamics represented in the model. While the first approach can be used to replicate a given time series (and, hence, replicate the exact inter-message arrival intervals), it does so by eliminating one of the most interesting source of variability in the model—namely the changing dynamics of the information flow. But if the modeler is to treat the information flow as an outcome of the model (rather than as an input) then the question becomes one of representing the right dynamics.

In our first attempt to model the live exercise that produced the chat data above, we hoped that the structural complexity of the command and control cell—64 operators in a many-to-many mapping to 15 chat rooms—would induce the “bursty” communication patterns. By representing the individual operators, each with a unique probability of replying to an incoming chat message, we expected that over time cascades of chat messages would emerge separated by long quiescent periods. While the model did just that, as Figure 3 shows, the rank-frequency plot of the simulated inter-message arrival intervals (shown in Figure 4) did not closely resemble the empirical data.

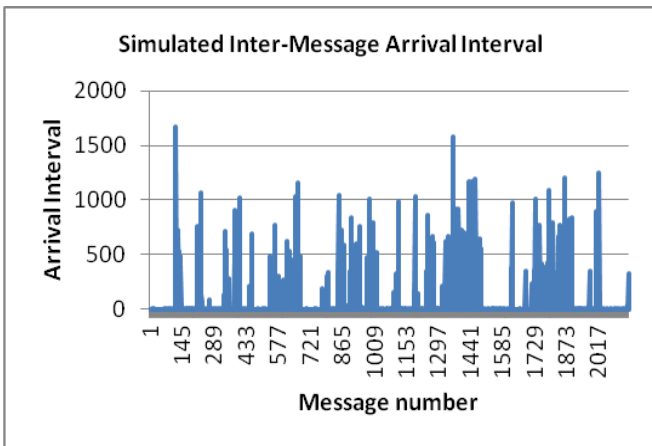


Figure 3 A C3TRACE simulation of inter-message arrival intervals. The graph depicts the time between consecutive messages.

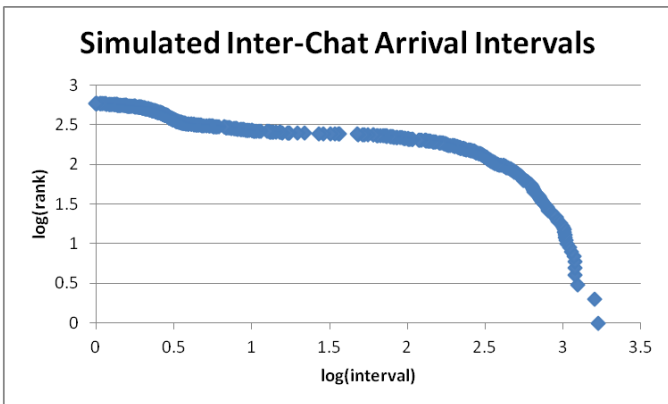


Figure 4 Rank-frequency plot of the inter-chat arrival intervals. The majority of the simulated inter-chat arrival intervals are well below a second and below the one-second resolution of the empirical data. We have not plotted the intervals of less than one second here.

It was clear that simply recapitulating the structure of the chat rooms would not by itself engender power law behavior. Similarly, changing the mean task times and their variability did nothing to change the shape of the interval distribution. In short, the complexity of the model itself was not enough to simulate the complex dynamics of communicative behavior.

For this reason, we returned to the literature to identify a more fundamental mechanism. Barabasi (2005) suggests that human communications patterns might be caused by a

“queuing process” in which the human must continually decide which message to respond to next. The idea is that the human associates a priority with each message and, given a choice among messages on a to-do list, will likely respond to the highest priority message first. In fact, these complex dynamics will be present even if the human must only decide between answering a given message and doing something (i.e., anything) else (Vazquez, 2005).

This minimal, two-task case is useful to the task network modeler insofar as it provides a template for representing dynamics where the salient inter-event intervals are task-based rather than queue-based. (In fact, it is not easy to represent a queue in which the priority-based sorting must be conditioned by a probabilistic threshold.) In the context of C3TRACE, this means we can generalize beyond the basic functionality of its priority-based message queues to represent a very generic and easily extensible priority-based task switching. Figure 5 depicts the task network diagram of a single operator’s priority-based selection between two tasks.

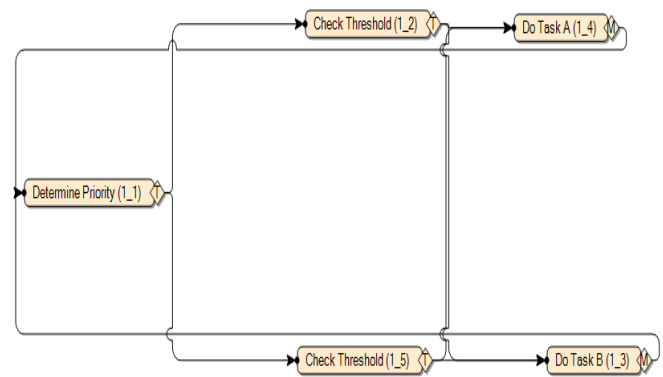


Figure 5 A generic two-task instantiation of Barabasi’s Priority List model in C3TRACE.

There are two important features of the model: First, each instance of the two tasks (“A” and “B”) is initially assigned a real-valued priority between zero and one and a new priority is assigned only after a task has been executed. Thus a task with low-priority value is likely to result in a relatively longer wait time while multiple instances of the tasks with higher priorities are processed. Second, there is a threshold value, independent of the priority value, that determines whether the higher-priority task will, in fact, be executed. A threshold value of one entails that highest priority task is *always* done first while a threshold value of zero entails random selection between the two tasks. As the rank-frequency plots in Figure 6 show, varying the threshold between these two extremes changes the likelihood of encountering extremely long wait times (even as the relative proportion of short wait times increases).

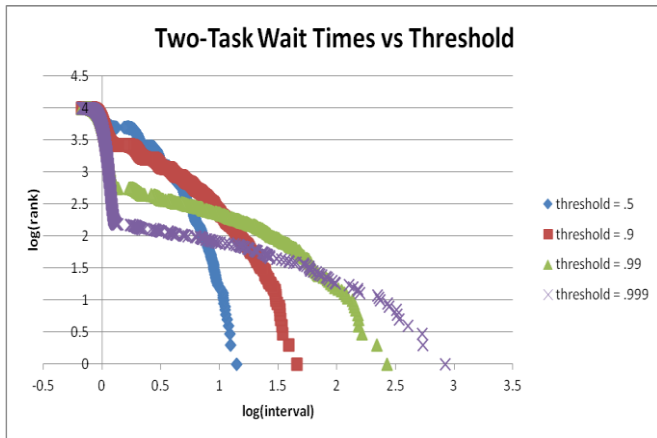


Figure 6 Rank-frequency plots of the wait times in the two-task model given different threshold values. The plots mirror those of Vazquez (2005).

The two task model above supports a straightforward extension to any fixed number of tasks and, unlike a queue-based representation, does not depend on the representation of additional constraints to ensure that the queue remains exactly saturated to engender the right dynamics (Barabasi, 2005). The decision whether to represent human communicative behavior directly by way of a priority-based queue or more abstractly by way of priority-based task switching will depend on the preferences of the modeler, but the ability to generalize the task-switching approach to other human behaviors weighs in favor of the task-based approach.

VALIDATING BEHAVIOR

While it remains an open question whether “retro-fitting” a complex model with the simpler task-switching approach just described will produce the right kinds of communication patterns, there is a more pressing concern. Namely, how do we determine if the underlying process is truly governed by a power law and, if so, how do we to identify the form of the power law distribution. Clauset, Shalizi and Newman (2009) have argued that despite its widespread use, identifying and fitting power-law behavior on the basis of least squares regression on log-log plots is fraught with problems. Instead, they argue that power law exponents should be derived using a maximum likelihood estimator (MLE) before a confidence interval is calculated indicating how likely it is that the data would be better fit by a “random” power-law process.

Estimating the smallest value (we refer to this as “ $x_{\min}$ ” below) for which a power law distribution holds is the first step in estimating the scaling index (i.e., the exponent) for the power law. Clauset et al. (2009) describe an iterative process for estimating that value. Each iteration starts with an estimate of $x_{\min}$ which is used to compute the complementary cumulative distribution function (CCDF) of the empirical data. The CCDF is determined by sorting the values from greatest to least and then simply computing the (complementary) percentile rank of each data point by dividing its rank in the sorted list by the total number of data points greater or equal to the estimated $x_{\min}$ value. We then estimate the scaling

index of the empirical data given our initial estimate of $x_{\min}$ using the MLE techniques (Newman, 2005; Clauset et al., 2009). Next, using the estimates for both $x_{\min}$ and scaling index, we compute the “theoretical” CCDF given the analytical solutions provided by Newman (2005). Finally, we determine the “distance” between these two CCDFs using the Kolmogorov-Smirnov (KS) statistic. In our cases, this distance is simply the greatest difference between percentile values of the empirical and theoretical CCDFs. The process repeats with each change in our estimate of $x_{\min}$ with the best estimate being the one that gives the lowest value (i.e., the least greatest value) for the KS-statistic.

Clauset et al. (2009) also describe a method for generating a p-value for a power law fit to the data. The basic idea is to compare the fit of the empirical data to the fit that would be obtained with a true power law distribution, using the same MLE techniques. More precisely, using the values for $x_{\min}$ and scaling index estimated from the power-law fit to the empirical data, we generate a synthetic set of data of the same size as the empirical data set. The values less than $x_{\min}$ are sampled directly from the same region in the empirical data. The values greater than $x_{\min}$ are generated using the estimated scaling index and the transformation method (Newman, 2005). This produces a synthetic data set with a true power law distribution above $x_{\min}$ and a distribution that differs from the power law in the exact way that the empirical data does below $x_{\min}$. We repeat this process, generating many synthetic data sets and fitting a power law to each one, using the same MLE techniques described previously. If the empirical data set is plausibly power law in the region above $x_{\min}$, then its KS-statistics should be in the same mix as those of the synthetic data sets. If the KS-distance is much larger than the range of synthetic KS-distances (i.e., its value falls in the tails), this indicates that the power law is not a good match to the empirical data. The proportion of synthetic data sets with KS-distances greater than that of the empirical data is the p-value. The lower the p-value, the less plausible it is that the empirical data are well-described by a power law. Higher p-values (above 0.1) mean a power law is fairly plausible.

So, for example, looking at the empirical chat data, we estimate an $x_{\min}$ of 84 seconds and power-law exponent of ~ 2.46 and a p-value of .65. Moreover, a similar fitting process to an exponential distribution yields a p-value of 0. Not only does this give us some reason to believe that the empirical chat data are the product of a power-law process, it also suggests a new approach for validating simulation data. Namely, assuming we have some confidence that both the empirical and simulated data are the result of power-law processes, we then simply compare the minimum values and power-law exponents.

CONCLUSIONS

Having implemented the methods advocated by Clauset et al. in C3TRACE, we have grown more confident that the human communication data discussed above are, in fact, the result of a power-law process, but this experience has taught

us several important lessons. (Our initial estimates of an $x_{\min}$ of 84 seconds and power-law exponent of ~ 2.46 produce a p-value of .65. Moreover, a similar fitting process to an exponential distribution yields a p-value of 0. While this suggests that a power-law fit is plausible, these results are only preliminary and a more thorough-going analysis is needed.)

First, even the most seemingly basic dynamics of human performance in an organizational structure are likely to reveal “non-normal” statistics. But these are exactly the kinds of dependent measures we need to represent

NO. OF
COPIES ORGANIZATION

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 DIRECTOR
(PDF) US ARMY RESEARCH LAB
IMAL HRA

1 DIRECTOR
(PDF) US ARMY RESEARCH LAB
RDRL CIO LL

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

1 ARMY RSCH LABORATORY – HRED
(PDF) RDRL HR

NO. OF
COPIES ORGANIZATION

8 ARMY RSCH LABORATORY – HRED
(PDF) SIMULATION & TRAINING
TECHNOLOGY CENTER
RDRL HRT COL M CLARKE
RDRL HRT I MARTINEZ
RDRL HRT T R SOTTILARE
RDRL HRT B N FINKELSTEIN
RDRL HRT G A RODRIGUEZ
RDRL HRT I J HART
RDRL HRT M C METEVIER
RDRL HRT S B PETTIT
12423 RESEARCH PARKWAY
ORLANDO FL 32826

1 ARMY RSCH LABORATORY – HRED
(PDF) HQ USASOC
RDRL HRM CN R SPENCER
BLDG E2929 DESERT STORM DRIVE
FORT BRAGG NC 28310

1 ARMY G1
(PDF) DAPE MR B KNAPP
300 ARMY PENTAGON RM 2C489
WASHINGTON DC 20310-0300

ABERDEEN PROVING GROUND

15 DIR USARL
(PDF) RDRL HR
L ALLENDER
P FRANASZCZUK
C COSENZO
RDRL HRM
P SAVAGE-KNEPSHIELD
RDRL HRM AL
C PAULILLO
RDRL HRM B
C SAMMS
RDRL HRM C
L GARRETT
RDRL HRS
J LOCKETT
RDRL HRS B
M LAFIANDRA
RDRL HRS C
K MCDOWELL
RDRL HRS D
B AMREIN
RDRL HRS E
N BUCHLER
D HEADLEY
L MARUSICH
W WARWICK

INTENTIONALLY LEFT BLANK.