Red Teaming Experiments with Deception Technologies

Fred Cohen, Irwin Marin, Jeanne Sappington, Corbin Stewart, and Eric Thomas

Abstract

This paper overviews a series of 30 experimental runs designed to measure the effects of deception defenses on attacks against computer systems and networks.

Background, Introduction, and Overview

As part of an overall effort to understand the implications of technical deceptions in information protection, an effort was undertaken to perform experimental assessment of the use of specific deceptive methods against human attackers. This effort represents only a beginning down the path of understanding the role of deception in information protection. [1] The specific set of technologies under study in this investigation were technologies similar to those described in earlier papers. [2]

Because of the high cost in time and material of such a study, many goals were tied to this effort. They included: (1) improving the understanding of the participants in how systems are attacked and how they can be defended, (2) understanding how much an attacker can be told about a deceptive defense before they are able to defeat it, (3) understanding how deception impacts attacker workload, (4) understanding the group dynamics underlying attack groups and how it relates to success and failure, (5) understanding what sorts of ideas, strategies, and tactics arise in such groups when they are not trained in any particular methodology of attack, and (6) understanding the impacts of initial access on the utility o deceptive defenses.

In total, 5 experimental runs of duration 4 hours each were run on each of 6 exercises. This represents 30 runs, including deception "on" and deception "off" control groups (6 each) and random "on" "off" mixes (18). Each run was preceded by a standard briefing and a run-specific briefing and followed by filling out of standard assessment forms, both individually by all team members and as a group. The exercises were of increasing intensity and difficulty so as to keep the participants challenged. Feedback was provided in the form of the exercise-specific briefing and was designed to first calibrate then systematically inform the attackers about more and more of the deceptive nature and type of the defense through the provision of 'intelligence' information being gathered by an insider. Eventually 'insider' access was granted to the attackers for measuring how they were able to perform with detailed knowledge and insider access to the nature of the deceptions. All experiments were repeated in very nearly identical circumstances with different groups of increasing suspected skill level and can bre repeated again in separate runs for other groups. A few of these experiments were repeated with higher quality attack groups with.

The Laboratory Environment

The laboratory environment used for these red teaming experiments consisted of two rooms.

- The first room is used by the attackers for their attacks. It consists of a set of attack computers and research computers. (1) The research computers are designed to provide the attackers with access to Internet and previously prepared capabilities and techniques as well as to provide access to additional computing capabilities, databases, and other individuals they may wish to seek help from. (2) The attack computers are configured in known configurations and are designed to facilitate attacks of the sorts known to the attackers. The attackers are permitted to, and often do, bring their own system capabilities to the exercise Systems in this room are instrumented to allow attack methods to be reviewed later and the room has a videotape machine for taping sessions. It also has a computer used by the observer to take

notes, is separated from the rest of the laboratory, and allows external access for bathrooms and other needs.

- The second room houses the systems under attack. It is physically separated from the attack room and is locked to prevent attackers from accessing it. It includes a set of systems and wiring capabilities to allow any network containing less than a few dozen computers to be rapidly configured and reconfigured to facilitate experiments.

The cost to supply such a laboratory is on the order of $40,000, most of which is in the cost of equipment. It took on the order of 50 person days to create the environment, In the case of these experiments, the laboratory itself is reasonably physically secure and has additional protections in this form of digital diodes to assure that information from experiments does not leak to the rest of the world. This is intended to assure that attacks do not spill over into the general Internet. A reasonable estimate of the costs of repeating these experiments would have to include the cost of labor (6 people for 5 hours for each run plus analytical time and experimental design and configuration time, and other support) comes to approximately $1000 per experiment plus $1,000 per run, or about $54,000 for this set of experiments. Facility space, electrical power, and other overhead bring the total cost of such an experiment to something on the order of $150,000.

Repeatability in Experiments

In order to assure essentially repeatable experiments, there are a set of file servers used to store complete disk images of experimental configurations. Using the Samba protocol and a bootable CD-ROM, we are able to make forensically sound images of systems to be attacked and systems launching attacks before and after experiments. The pre-experiment images are reloaded into experimental systems prior to each experiment so that all systems involved in the experiment are, in essence, identical. The one exception is that experiments are run on different days, and sometimes at different times of day to accommodate schedules.

The ability to create a very nearly identical experimental environment is critical to such research and there is a considerable cost associated with this. For example, even at relatively high network speeds, it costs on the order of 12 minutes per system to make an image and another 12 minutes to restore that image. This means that reproducing an experiment requires something like an hour of preparation time as well as possible network reconfiguration.

All experiments are permanently archived so that they can be repeated at a later date and time by the same group or another group of test subjects. This allows effects like training, experiment order, and subject biases to be remediated and allows groups to repeat experiments after training, after being provided with additional information, and after intentional introduction of biases.

Effects Under Consideration

In the initial 45 experiments performed in this environment we were most interested in several primary factors:

- The difference in performance with and without deceptions in place is fundamental to our desired understanding. In order to observe this effect, open ended exercises are used. In these sorts of efforts, the problem is sufficiently complex for the time provided that it would be an exceptional team that could complete all facets of the challenge in the allotted time. The experiments have sets of goals that, in essence, require the achievement of some earlier objectives to achieve some later objectives. The objective of deceptions in this case is to reduce the effectiveness of attackers. The metric is then how far the attackers get how fast rather than their ability to complete all tasks. In this sense, the problems are like mazes without end and the characterization we use to describe them later is an attack graph. A fully

successful attack would, presumably, have to follow one of a small number of attack graphs that lead to success. Other graphs lead to false success (when deception is in place) or to failures or delays. We can then measure success relative to finding one of the paths that leads down a successful attack graph.

- The difference in performance of attackers between situations when the deception is known to the attacker and when it is unknown to the attacker was also vital to our understanding because we were interested in the performance of deceptive defenses in the presence of insider threats, intelligence threats, and overrun threats. Thus we performed experiments with different levels of knowledge provided to the attackers so that we could measure the performance difference based on their knowledge of the situation.
- Based on some initial theoretical work we believe that there may be a correlation between success and the types of deceptions we are trying to induce. Specifically, we sought to differentiate deceptions that induce type 1 (omission), 2 (commission), and 3 (misdirection) errors and to understand he thresholds at which these types of errors occur, are detected or suspected by attackers. and can be induced with effect.

In this initial experiments, only these three factors were explored, however, we are also interested in aspects of the nature of deception [1] and the way in which they operate in the information defense arena. Specifically, we are interested in how limited resources lead to controlled focus of attention, how effective deceptions can be composed from concealments and simulations, how memory and cognitive structure force uncertainty, predictability, and novelty and how this can be exploited for deception, how time, timing, and sequence work in deceptions, how much control over observables are required, operational security requirements, effects of different attack methodologies and capabilities, the recursive nature of deceptions, how small changes can impact large systems, the complexity required for implementing deceptions to great effect, what level of knowledge of the target is required to be effective over what time frames, how deceptions can be modeled and outcomes predicted, and how counterdepcetion functions.

Additional Goals of Exercises

As part of these exercises, we also hoped to advance the knowledge and skills of the participants. The participants, in this case, were students ranging in age from 16 to 38, all in computer-related fields, all with excellent grade point averages, all US citizens, and all interested in information protection, and all participating in an intensive program of study and research in this area. Through this effort, we hoped to give them skills and knowledge that would be helpful in understanding how systems are attacked and how they may be more effectively protected. The students were also taught classes on information protection, received training in how to manage and operate systems, and participated in hands on research and systems administration projects over the period of this effort.

The same exercises were also run on more skilled attackers including teams of professionals that do testing of high assurance systems, professional red teaming groups, professionals in the field of information system intelligence, and professional offensive information warriors. These experiments are used to calibrate the results. This paper does not include these results in its findings because they were not statistically meaningful, however, those results were consistent with what we found for the sample group under study.

Summary of Collected Data

The collected data consists of evaluation forms filled out by all participants after each session, a group form filled out as a consensus in a facilitated group meeting after individual forms were completed, a summary of events and times as recorded by the observer, and detailed copies of the

system configurations before and after each exercise [3] Standard pre-briefings were provided for each group to assure to a reasonable extent that groups would keep results independent of each other and to provide reasonable limits on behavior while fulfilling administrative requirements of the facility. [5] Forms were designed so as to solicit specific information related to research interests. [4] Specifically, questions were directed toward determining whether deceptions were thought to have been identified and bypassed, understanding whether the participants were operating in level 1, 2, or 3 of the cognitive characterization used in the framework for deception [1] which forms the basis for this work, detecting issues in group behavior that relate to success and failure of deceptions (e.g., the effect of the group on preventing exploration of lines and the effect of the group on inducing lines), and information on the strategies employed and tools use and effectiveness, which are directed at improving performance of other groups in similar tasks.

After each set of experiments, full details were provided to all participants. Thus the sequencing of experiments went from (1) no revelation of deception issues to (2) provision of details about the presence of deceptions and the deception technologies in use and finally (3) to full details of the deceptions including all configuration details. This enabled us to measure across the dimension of knowledge of the deception. Control groups were used with deception always off and deception always on so that cross-experiment differences in time to achieve goals could be measured. These groups were maintained within each sub experiment (3 weeks duration) but groups were reshuffled after each three week period to try to find group mixes that tended to improve performance on red teaming efforts and to help students learn how to work well in project groups and learn more from each others talents and skills.

There were also faults detected in experiments. While we do not believe that any of these faults invalidate the overall results, additional experiments and improved experimental conditions would be helpful in mitigating such faults in the future. Specifically, fault fell into the following categories; (1) limits of the facilities and situation, (2) limits of the experimenters and time frames, (3) limits of the technology employed, (4) experimenter and participant error.

- **limits of the facilities and situation** The facilities were being upgraded and altered under us while these experiments were being performed and the facility was never intended for this sort of experiment. Interruptions were kept to a minimum, but they did occur, a network outage interrupted the location of Internet data on one occasion, the technology used to facilitate the work was less than ideal, and there were days without air conditioning when it was over 80 degrees Fahrenheit in the attacker's facility. We did all we could to keep things equitable, but clearly these sorts of conditions have some impacts on performance.
- **limits of the experimenters and time frames** The experimenters involved were not professionals in this realm and thus were not perhaps as good as their jobs as some others might not have been. In addition, it was necessary for the observer to have knowledge of the real situation and to be in the same room as the subjects. Thus there was the potential for bias and, on some occasions, there was laughing by observers and interaction between subjects and observers. The time frames for setting up and running these experiments were also very tight, so experiments did not always function perfectly and imperfections observed by the observer were repaired while the experiment was ongoing. while efforts were made to avoid any direct information from this activity, on several occasions subjects suspected that the observer had altered the experiment.
- **limits of the technology employed** The specific deception technologies employed were thrown together on very little notice, as was the environment for the deceptions. This was because of the short window of opportunity to collect data while there were enough subjects

available. This caused a variety of complexities, but for the most part, the same conditions were present for each group so that these issues tended to even themselves out.

- **experimenter and subject error** These included cases where experimenters and participants made various mistakes. In particular: (1) We had two cases where a subject indicated that they had reversed the meanings of the numerical values in the evaluation forms during the out briefing when all participants were asked to come up with numerical values together. We corrected the values in these subject's forms immediately thereafter by inverting the values (5 became 1, 1 became 5, and so forth). (2) In one experiment an error in system configuration prohibited progress for more than an hour. This was mitigated during the experiment and the time difference between the time the same activity that showed the error and the time when it was compensated for was subtracted from subsequent times in the results. (3) In a few cases the familiarity of the subjects with the observers, the presence of additional observers, or the presence of a camera in the room caused limited interference with the experiments, however, we do not believe that these had any effects on the progress relative to the attack graph from a standpoint of differences between the presence and absence of deceptions. Specific cases are noted below where appropriate.

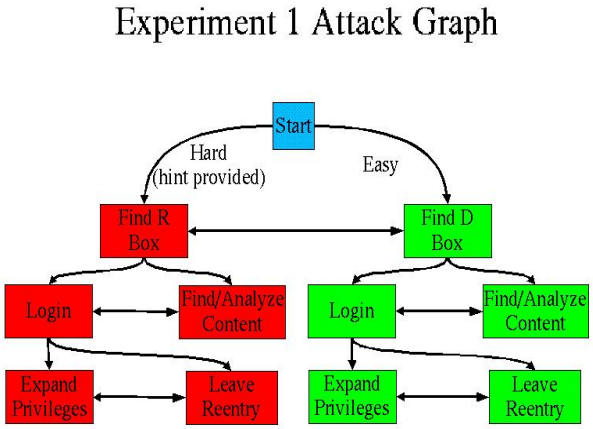
Finally, as in many such experiments, the subjects were predominantly academically skilled college students studying computer security at a national laboratory. While these results look promising, such students almost certainly represent only a small segment of the space of real attackers, and are far less skilled than many real attackers. Select experiments were also performed with other groups and details are provided for those cases below.

It would clearly be desirable to repeat these experiments under more realistic conditions, however, we do not believe that these conditions had any serious impact on outcomes and we believe that money spend on such efforts would be better spent doing other experiments which provide additional results while covering the issues in this set of experiments as a side effect of those efforts to detect any refutations should they arise, or to provide confirmations of these results.

The Structure of Attack Graphs

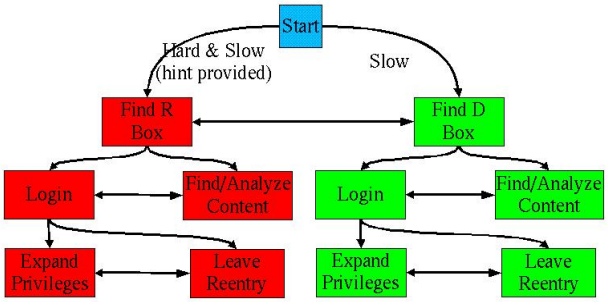
In each experiment, there were known successful attack graphs and actual attack graphs followed by participants. In this section, we summarize the successful attack graphs for each run, so that they can be compared to actual attack graphs, and alternative attack graphs yielding type 1, 2, and 3 errors, as observed in experiments. Unlimited numbers of additional attack graphs are likely feasible for successful attack, seemingly successful attack (deceptions effective), and failed attacks.

1R or 1D find box (easy) D directs target to wrong victim
2R or 2D log in | find content (Wrong path looks good) find content | analyze content (Wrong path looks good) analyze content | login (Wrong path looks good)
3R or 3D leave reentry | expand privileges (Wrong path looks good) expand privileges | leave reentry
4R or 4D target believes they win when they lose and deceiver observes and learns about target



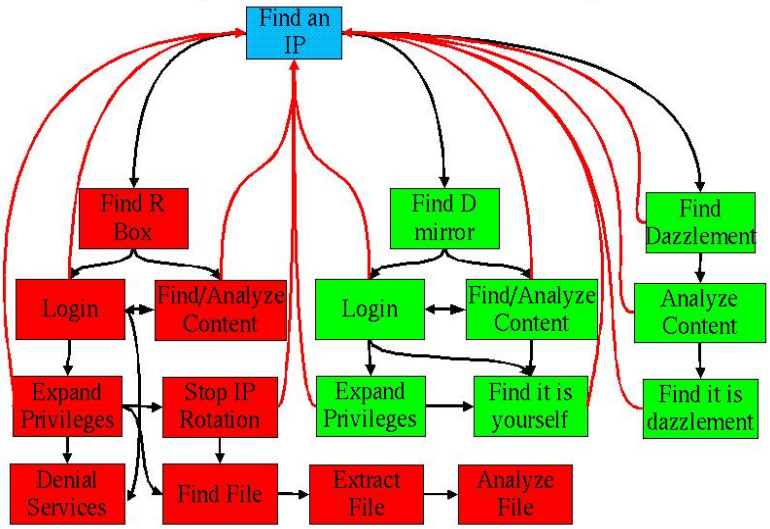
1R or 1D find box (hard) - D directs target to wrong victim, search is very slow, time pressure induces alternative search strategies, some search strategies reveal deception - but are not noticed
2R or 2D log in | find content (Wrong path looks good)
find content | analyze content (Wrong path looks good)
analyze content | login (Wrong path looks good)
3R or 3D leave reentry | expand privileges (Wrong path looks good)
expand privileges | leave reentry
4R or 4D target believes they win when they lose and deceiver observes and learns about target

Experiment 2 Attack Graph



1R or 1D loop: find box - Deception makes differentiating box harder and increases find (real) box time dramatically
2R or 2D log in | find/analyze content (Wrong path consumes time)
Addresses change before success => goto loop
Trigger detector => goto loop w/shorter times
3R or 3D time low => deny services - but deny to what? - and tell how?
4R or 4D leave reentry | expand privileges
leave bug => easier to find
5R stop movement => easier to find | plant Trojan => easier to find
6R find file
7R extract file and analyze file

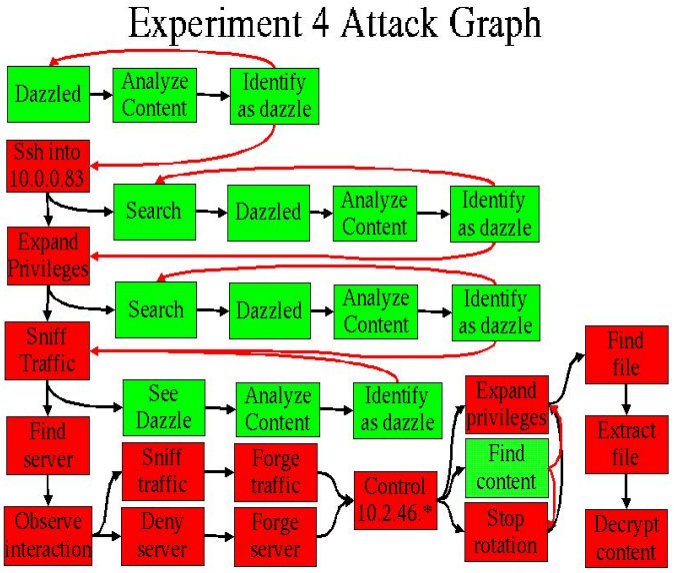
Experiment 3 Attack Graph



1D Search for or try to analyze 10.0.0.83 and ignore intelligence provided

1R Enter 10.0.0.83 via ssh
2D Search for other systems in 10.0.*.* and try to exploit them
2R Expand privileges using routine provided
3D Search network for targets to attack
3R Sniff traffic
4D See dazzlement, analyze, identify as dazzlement
4R Find real client and server and observe traffic
5R Understand interaction and determine a viable attack
6R Gain control of the victim
7D Look for content (unfindable in this state)
7R Expand privileges
8R Find file
9R Extract file and analyze content

Experiment 4 Attack Graph



While these high level representations of attack graphs are not strictly accurate to the details of attack sequences, they are helpful in understanding the nature of the situation. Metrics could reasonably be related to each link in these graphs with the resulting weighted graph providing measures for the difficulty of attack given the deception situation. Creating these weights requires two things. (1) There are strictly mathematical issues, such as the number of paths in some direction and their distribution, that might lead to purely mathematical values for some metrics. For these direct solutions can be applied. (2) The rest of the situations depend on the relative skill of the attacker in detecting the victim and differentiating it from the deceptions. This detection and differentiation problem comes down to peoples' ability to devise automation and use their own analytical capabilities. This sort of data can only be found through empirical measurement, or in other words, experiments.

The use of attack graphs such as these provide a means for measuring the effectiveness of defenses by measuring the progress by attackers over time. It can be reasonably argued that the slower an attacker moves toward the objective of their attack the better the defender is at defending against the attack. To the extent that the attacker moves down a deception line, they can be thought of as moving further from the target of their attack. This brings up the notion of plotting progress over time with progress toward deceptions as negative and progress toward real targets as positive. While such measurements are not in proportion in the sense that a positive 2 is twice as far as a positive 1, they do form a POset so that a position of n is always better than a position of n-1 for the attacker.

It is also important to note that these cases are all examples of source-target (s-t) graphs and that there in an underlying assumption that the attacker is goal directed toward getting to a target from a source and the defender is seeking to prevent or delay the attacker from reaching the target.

Actual Graphs Followed

Each group in each experiment followed an actual attack graph over time. These attack graphs are summarized here along with some interpretation. We use the term "Hop" interchangeably with "Experiment" and indicate the first time (T) the attacker got to any given step (S) (in the case of some deception systems, steps may have to be retried many times).

Group	***Hop***	***D***	***S/T***	***S/T***	***S/T***	***S/T***	***S/T***	***S/T***	***S/T***
Mon	1	No	1R 2:00	2R 2:08	3R 2:45				
Tue	1	No	1R 0:22	2R 0:24	3R 1:11	4R 3:27			
Wed	1	No	1R 1:58	2R 1:58					
Thu	1	Yes	1D 0:17	2D 0:20	3D 0:22	4D 2:26			
Fri	1	Yes	1D 0:31	2D 0:31	3D 3:08	4D 3:23			
Mon	2	Yes	1D 3:37						
Tue	2	No	1R 3:33						
Wed	2	No	1R 1:37	2R 1:42					
Thu	2	Yes	1D 1:48*	2D 2:06					
Fri	2	No	1R 0:40	2R 0:49					
Mon	3	No	1R 0:41	2R 1:25					
Tue	3	No	1R 1:15	2R 2:58					
Wed	3	Yes	1D/R 0:52						
Thu	3	Yes	1D/R 0:17						
Fri	3	No	1R 0:29	2R 0:51					

Group	*Hop*	*D*	*S/T*	*S/T*	*S/T*	*S/T*	*S/T*	*S/T*	*S/T*
Mon@	4-1	Yes	1D 0:38	1R 2:07	2R 2:16	2D 3:01	3D 3:20		
Tue	4-1	No	1D 0:30 *+1*	1R 0:45 *+2*	2D 0:50	2R 1:40	3R 1:45	4R 1:50	
Wed	4-1	No	1R 0:21	1D 0:30	2R 0:42	3R 1:05	4R 1:30	5R 2:45 *+3*	
ThuA	4-1	Yes	1R 1:34	2R 1:45					
Thu	4-1	Yes	1D 0:55	1R 1:35	2R 1:50	3D 2:23	1D 2:23 *+4*	3D 2:55	
Fri	4-1	Yes	1R 0:37	2R 0:54	3D 1:43	1D 2:31	4D 2:43	3R 3:37	
Mon+	4-2	Yes	1D 0:51	1R 1:32	2R 1:41	3R 1:45	4D 1:45		
Tue+	4-2	No	1D 0:34	1R 1:22	2R 1:33	3R 2:10	4R 2:10		
Wed+	4-2	No	1R 1:45	2R 2:18	3R 2:30	4D 3:12 *+5*			
Thu+	4-2	Yes	1R 0:47	2R 0:58	3D 1:12	3R 3:15			
Fri+	4-2	Yes	-	-	-	-	-	-	-
Mon+	4-3	Yes	1R 0:20	3R 0:59	2D 1:45	2R 1:59	3R 2:06	3D 2:22	4R 3:01
Tue+	4-3	No	1R 0:27	2R 0:28	3R 1:10	4R 1:24			
Wed+	4-3	No	1R 0:18	2R 0:19	3R 0:23	4R 1:32	5R 3:10		
Thu+	4-3	Yes	-	-	-	-	-	-	-
Fri+	4-3	Yes	-	-	-	-	-	-	-
SR-1+*6*	3.1	Yes	1D/1R	2D/2R	3D/3R				

* They achieved 1R at 2:06 but never realized it because they were occupied with following the line of 2D.

@ Groups were realigned after the third run to meet changing schedules and to allow the groups to team with those they thought they would work together with best. Several teams stayed together, one participant opted out of the process in favor of other work. In addition, after the third set of experiments all teams were briefed out on the deceptions that were in use, how they might have succeeded, and provided with full details of the technologies in use and how they worked. This included the provision of access to source code for the deception technologies under study.

+1 Even with deception turned off, teams try various lines that are not fruitful. They did not observe a deception, which accounts for rapidly moving to 1R. On the previous day, the deception caused about 1.5 hours of delay.

+2 Due to an experimental fault 1:45 was wasted between 0:30 and 0:45, so times have been adjusted backwards to reflect progress toward the goal.

+ Experiment 4 was run three times on the same groups to give them more opportunity to spend more time on the same problem, including the development of improved tools.

+3 They see the interaction but do not yet realize what it really is.

A On this particular Thursday we had an additional exercise in the morning (AM) that ran for 4 hours and involved the team that designed the experiments (but not the person who built the specifics of this run).

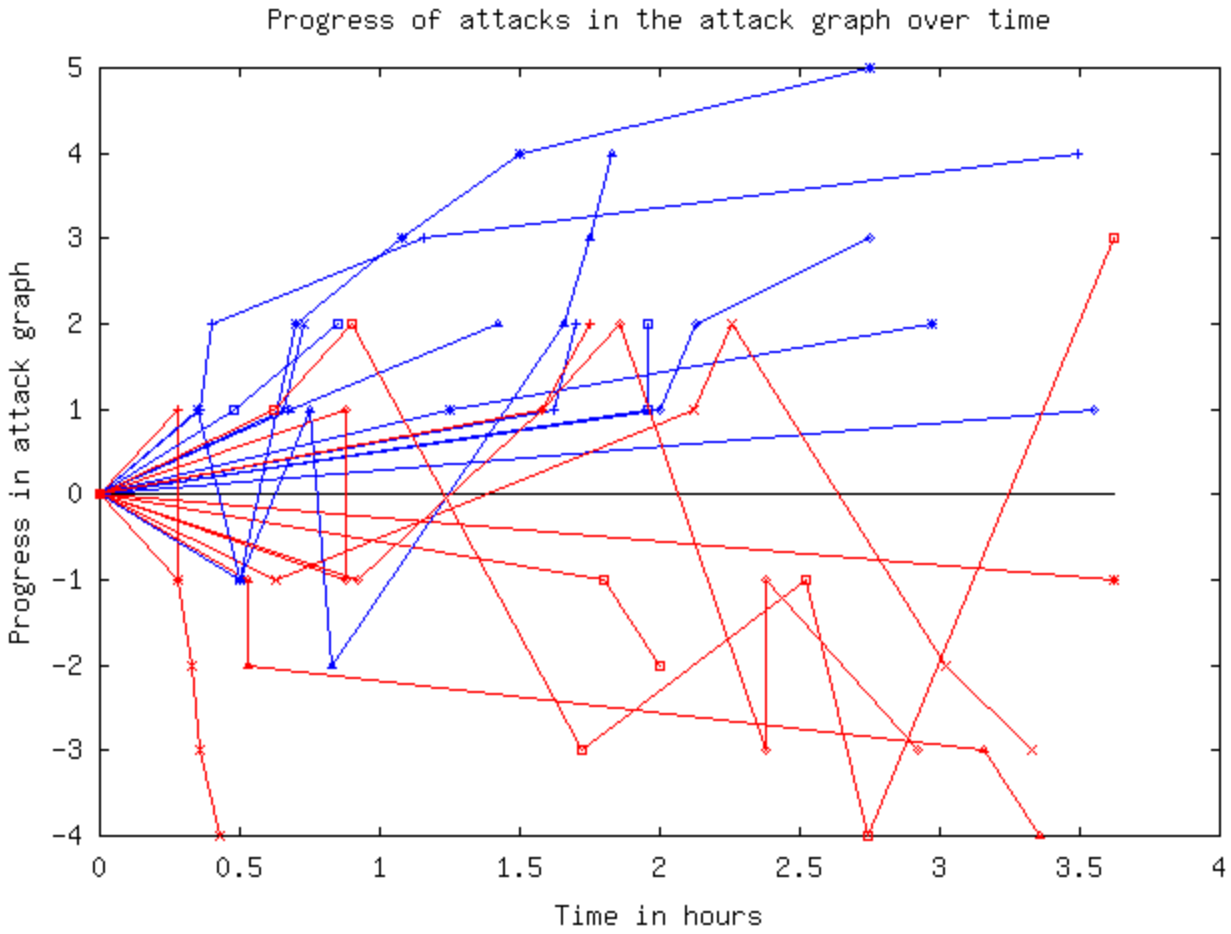
+4 They lose confidence in the real line because of dazzlements (3D) and return to 1D believing the original dazzlement over the real system they were in.

+5 they do not differentiate their own scans and deceive themselves temporarily.

+6 This was an 'extra run' of a slightly enhanced experiment 3. Details are provided below under 'special runs'.

- indicates a team that decided not to participate.

The following plot summarizes this data in a different format. In this summary, each run is represented by a line. Lines in red indicate attack sequences with deception enabled while lines in blue show attack sequences with deception disabled. The 'X' axis represents time, while the 'Y' axis is positive for 'Real' locations in the attack graph and negative for 'Deception' locations in the attack graph. These plots show the progress toward the target from the source as a function of time.



If anything is clear from this plot it is that attackers do better without deception. This is no surprise. However, there are a lot of other interesting characteristics in these results that we will now discuss.

The following table summarizes detailed information on factors identified for measurement in the experiment and called out in the provided forms. The data fields below comprise numerical responses to the following question areas: Date, Deception (Yes or No) Identification, Teamwork effectiveness, Strategy import, Strategy effectiveness, New strategy import, New strategy effectiveness, Extent of success, Import of success, Time pressure, Uncertainty, Distractions,

Exhaustion, Difficulty, Interest level, Enjoyability, and Surprise. Detailed questions are included in the "Red Teaming Questionnaire Form" [4] cited earlier.

Date	*D*	*ID*	*Team*	*SI*	*SW*	*NSI*	*NSW*	*Suc*	*ISuc*	*Time*	*Uncert*	*Dist*	*Tired*	*Hard*	*Int*	*Joy*	*Surp*
2001-06-04	N	JD	3	1	1	3	3	2	4	3	4	1	3	3	3	3	5
2001-06-04	N	JR	3	3	2	3	3	2	2	3	5	3	4	5	1	3	2
2001-06-04	N	SM	3	4	2	4	3	2	5	3	4	3	3	3	3	2	4
2001-06-05	N	JD	5	5	5	3	3	5	5	2	2	2	2	1	1	3	4
2001-06-05	N	OO	5	5	4	4	4	5	5	5	3	2	3	3	4	4	5
2001-06-05	N	MC	5	4	5	3	3	5	5	3	4	2	1	2	4	4	4
2001-06-06	N	MP	4	2	2	3	2	2	3	3	2	2	4	5	4	3	4
2001-06-06	N	CK	3	1	1	3	3	1	3	2	2	4	2	4	3	3	4
2001-06-06	N	JA	3	3	3	3	3	2	3	3	4	2	3	4	4	4	3
2001-06-06	N	GS	2	3	2	3	2	2	5	5	4	2	4	4	5	5	3
2001-06-07	Y	GG	4	3	4	3	3	3	3	2	3	2	2	3	2	2	2
2001-06-07	Y	RW	3	3	4	3	3	5	5	1	4	2	1	2	3	4	3
2001-06-07	Y	SD	4	3	3	5	3	4	5	1	4	2	4	1	2	3	4
2001-06-07	Y	JS	3	4	3	3	3	4	3	2	5	2	3	4	4	3	4
2001-06-08	Y	DH	2	3	3	3	5	4	3	1	4	2	2	3	3	3	5
2001-06-08	Y	AC	2	2	5	3	4	5	4	3	3	3	3	2	2	2	3
2001-06-08	Y	LD	3	2	5	3	3	2	3	5	4	3	2	4	5	4	3
2001-06-08	Y	LA	3	3	3	3	5	5	5	2	4	2	2	3	3	4	3
2001-06-11	Y	JD	2	3	1	3	1	1	5	1	1	1	4	5	3	3	5
2001-06-11	Y	SM	3	3	1	1	1	1	5	3	4	1	4	4	3	2	3
2001-06-11	Y	JR	1	1	2	1	2	3	3	4	1	4	2	4	2	3	2
2001-06-12	N	JD	4	3	3	3	3	3	3	3	4	3	3	5	3	3	3
2001-06-12	N	OO	3	3	2	3	3	2	3	3	3	3	4	4	4	3	3
2001-06-12	N	PS	3	3	3	3	3	3	3	3	3	3	3	5	5	4	4
2001-06-12	N	MC	3	3	2	3	3	2	5	4	5	4	2	5	3	3	4
2001-06-13	N	MP	3	4	4	3	3	3	5	2	3	3	3	5	4	3	3
2001-06-13	N	CK	3	2	1	2	1	1	4	3	4	4	3	4	3	2	3
2001-06-13	N	GS	3	5	2	3	2	1	5	5	3	3	5	5	5	5	3
2001-06-13	N	JA	3	3	2	1	1	2	3	2	3	2	4	4	3	3	3
2001-06-14	Y	GG	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2001-06-14	Y	RW	3	3	2	2	2	2	5	3	4	2	3	4	4	4	3
2001-06-14	Y	JS	4	3	4	3	3	3	3	3	2	4	3	4	4	4	3
2001-06-14	Y	SD	3	4	2	3	3	3	3	2	5	3	2	4	2	1	3
2001-06-15	N	AC	1	1	1	1	1	1	5	5	5	5	5	5	1	2	3
2001-06-15	N	DH	2	1	1	3	3	1	2	4	3	3	3	4	3	3	3
2001-06-15	N	LD	3	2	2	2	2	1	3	1	3	1	4	5	3	3	4
2001-06-15	N	LA	1	1	1	2	1	2	3	3	2	3	2	4	3	2	4
2001-06-15	N	BB	3	1	1	3	2	2	3	1	1	1	3	4	2	2	1
2001-06-15	N	CK	4	4	3	3	3	3	3	3	3	3	3	4	4	3	3
2001-06-18	N	SM	5	5	4	3	3	4	5	4	1	1	4	4	5	5	4
2001-06-18	N	JD	3	3	3	3	3	4	5	5	2	1	1	3	4	4	3
2001-06-18	N	JR	2	2	3	1	4	3	2	3	5	3	3	4	5	1	4
2001-06-19	N	JD	5	4	4	4	4	4	5	3	3	1	1	3	3	3	4
2001-06-19	N	OO	4	4	4	4	3	4	5	3	3	2	3	3	4	3	4
2001-06-19	N	PS	4	4	3	3	3	3	4	4	3	3	1	3	4	4	3

Date	*D*	*ID*	*Team*	*SI*	*SW*	*NSI*	*NSW*	*Suc*	*ISuc*	*Time*	*Uncert*	*Dist*	*Tired*	*Hard*	*Int*	*Joy*	*Surp*
2001-06-19	N	MC	5	4	5	4	5	4	5	5	2	1	1	2	4	4	5
2001-06-20	Y	AN	4	4	4	3	3	3	5	3	3	3	3	5	5	5	3
2001-06-20	Y	GS	4	4	4	3	3	3	5	3	3	3	3	4	5	3	5
2001-06-20	Y	JA	3	3	2	3	3	2	4	3	2	2	3	4	4	4	4
2001-06-20	Y	MP	4	4	2	3	3	2	4	3	3	2	3	5	5	4	4
2001-06-20	Y	CK	4	4	4	3	3	3	4	3	4	2	2	5	4	4	4
2001-06-21	Y	GG	4	2	3	3	3	4	3	3	3	3	3	3	3	5	3
2001-06-21	Y	RW	3	3	3	3	3	2	4	4	3	2	4	4	5	5	3
2001-06-21	Y	JS	4	3	1	3	3	2	3	4	3	1	2	4	5	4	3
2001-06-21	Y	SD	3	4	2	3	3	2	4	2	5	4	4	4	3	2	2
2001-06-21	Y	VN	4	3	3	3	3	2	3	2	4	1	2	5	5	5	3
2001-06-22	N	AC	3	1	3	3	3	1	5	4	2	5	3	5	3	3	3
2001-06-22	N	LD	1	1	1	1	1	2	3	2	1	5	5	5	4	3	3
2001-06-22	N	DH	3	2	1	5	1	2	1	2	2	5	2	4	5	3	4
2001-06-22	N	BB	3	3	3	3	3	3	3	2	3	1	3	3	4	4	3
2001-06-22	N	CK	3	1	1	1	1	2	4	3	5	5	5	4	4	2	3
2001-06-22	N	LA	2	2	1	3	3	1	1	1	2	2	3	4	3	2	4
2001-06-25	Y	SM	4	4	4	3	3	3	5	4	2	3	3	4	4	3	1
2001-06-25	Y	SD	4	3	2	3	3	2	3	3	3	2	2	4	4	4	3
2001-06-25	Y	BB	2	2	1	3	3	1	3	2	4	1	3	5	3	2	3
2001-06-25	Y	JD	3	1	1	1	1	1	3	3	3	1	3	5	3	2	3
2001-06-25	Y	JR	3	3	3	2	3	2	2	1	3	1	2	4	3	3	3
2001-06-25	Y	KM	3	1	1	1	1	1	5	5	5	2	2	5	5	5	3
2001-06-26	N	MP	3	4	3	3	3	3	3	3	4	2	3	4	4	3	4
2001-06-26	N	GS	2	2	2	3	3	2	4	4	4	2	4	5	3	3	3
2001-06-26	N	PS	3	3	3	3	3	3	4	4	4	4	3	4	3	3	3
2001-06-26	N	JA	3	4	4	3	3	4	4	4	3	2	2	4	5	4	4
2001-06-26	N	NP	4	4	2	4	4	2	3	4	3	3	5	5	4	3	2
2001-06-26	N	MC	5	5	4	3	3	4	5	4	4	4	1	4	3	4	4
2001-06-27	N	GG	4	3	3	4	3	4	3	3	3	3	3	3	3	3	3
2001-06-27	N	RW	3	5	2	3	2	2	4	3	3	2	1	5	5	3	4
2001-06-27	N	JS	5	4	3	4	3	4	3	2	3	2	1	3	5	5	3
2001-06-27	N	AN	4	3	3	3	3	3	5	3	4	5	3	5	4	4	3
2001-06-27	N	VN	5	5	5	3	3	3	3	1	4	3	3	4	3	3	2
2001-06-27	N	OO	3	5	3	4	3	2	5	4	3	3	3	3	3	3	3
2001-06-28	Y	CK	3	2	2	3	3	2	4	3	4	4	3	4	3	4	4
2001-06-28	Y	RY	4	1	1	1	1	2	2	2	4	3	3	4	5	3	4
2001-06-28	Y	BS	2	4	3	3	3	2	4	3	3	3	2	4	5	3	3
2001-06-28	Y	NB	3	3	3	3	2	3	3	3	3	3	2	4	5	4	3
2001-06-29	Y	JD	4	3	3	3	3	3	3	4	4	3	3	5	1	1	3
2001-06-29	Y	DH	3	3	3	3	3	3	3	1	3	2	3	4	3	3	3
2001-06-29	Y	LA	3	2	1	3	2	1	4	3	3	2	2	5	3	2	4
2001-06-29	Y	CK	3	3	3	2	3	3	4	2	4	3	3	3	5	3	5
2001-06-29	Y	LD	2	1	1	3	1	2	2	1	5	2	5	3	2	1	3
2001-07-09	Y	JD	2	1	1	1	1	1	5	1	4	3	3	5	3	2	3
2001-07-09	Y	JR	3	3	3	3	3	1	2	1	5	5	5	5	1	1	4
2001-07-09	Y	SM	3	3	2	3	3	3	3	3	3	5	5	5	3	1	3

Date	*D*	*ID*	*Team*	*SI*	*SW*	*NSI*	*NSW*	*Suc*	*ISuc*	*Time*	*Uncert*	*Dist*	*Tired*	*Hard*	*Int*	*Joy*	*Surp*
2001-07-09	Y	JC	3	5	5	5	2	3	5	1	4	1	4	4	4	4	5
2001-07-09	Y	BB	2	2	1	1	1	1	3	2	5	4	3	5	3	2	3
2001-07-09	Y	KM	2	2	1	1	1	1	3	1	4	4	5	4	1	2	4
2001-07-09	Y	SD	4	4	3	3	2	2	4	2	2	4	3	4	4	4	3
2001-07-10	N	NP	4	3	3	3	3	4	5	4	3	2	3	5	4	4	4
2001-07-10	N	JA	4	3	3	3	3	3	4	4	3	2	3	4	4	4	4
2001-07-10	N	GS	5	3	3	3	3	5	5	5	3	3	5	5	5	3	3
2001-07-10	N	PS	3	3	3	3	3	2	4	3	4	3	3	4	3	2	3
2001-07-10	N	MP	4	4	3	3	3	3	5	4	4	2	3	5	4	3	3
2001-07-10	N	MC	2	2	1	3	3	1	4	2	5	3	4	4	2	1	3
2001-07-11	N	GG	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2001-07-11	N	JS	2	3	2	3	2	2	3	2	4	3	3	5	2	2	2
2001-07-11	N	LA	3	3	3	3	3	2	3	3	1	2	4	5	2	1	1
2001-07-11	N	VN	3	3	3	3	3	1	3	1	5	3	3	5	2	2	3
2001-07-11	N	OO	4	4	3	1	1	3	4	2	3	3	3	4	3	3	3
2001-07-12	Y	NB	4	3	1	3	1	3	3	1	2	3	3	5	4	5	4
2001-07-12	Y	RY	3	3	2	3	3	1	4	2	4	4	3	5	3	3	3
2001-07-12	Y	CK	2	4	1	3	3	1	4	2	4	5	2	4	3	4	3
2001-07-12	Y	BS	4	3	1	3	3	1	3	3	3	2	3	5	4	3	3
2001-07-16	Y	BB	3	3	3	3	3	2	2	3	3	1	3	4	2	2	3
2001-07-16	Y	JD	5	5	5	3	3	3	3	1	3	1	1	5	5	5	3
2001-07-16	Y	SD	4	4	4	3	3	4	4	2	4	4	3	4	3	2	3
2001-07-17	N	NP	3	4	3	3	3	3	4	4	4	3	5	4	1	1	4
2001-07-17	N	JA	3	3	2	3	3	3	2	2	3	2	4	4	2	2	3
2001-07-17	N	PS	5	4	4	3	3	3	4	4	3	3	1	3	4	4	4
2001-07-17	N	MC	3	4	2	3	3	2	5	4	3	3	1	4	4	3	4
2001-07-17	N	GS	5	3	3	5	3	3	5	5	3	3	3	3	5	3	5
2001-07-18	N	VN	4	3	3	3	3	3	3	2	4	1	3	5	3	3	3
2001-07-18	N	AN	4	3	3	3	3	3	4	3	3	1	3	5	4	3	3
20007-18	N	OO	4	3	3	3	2	3	4	1	3	1	2	4	3	3	3

Data on Confounding Factors

Analysis

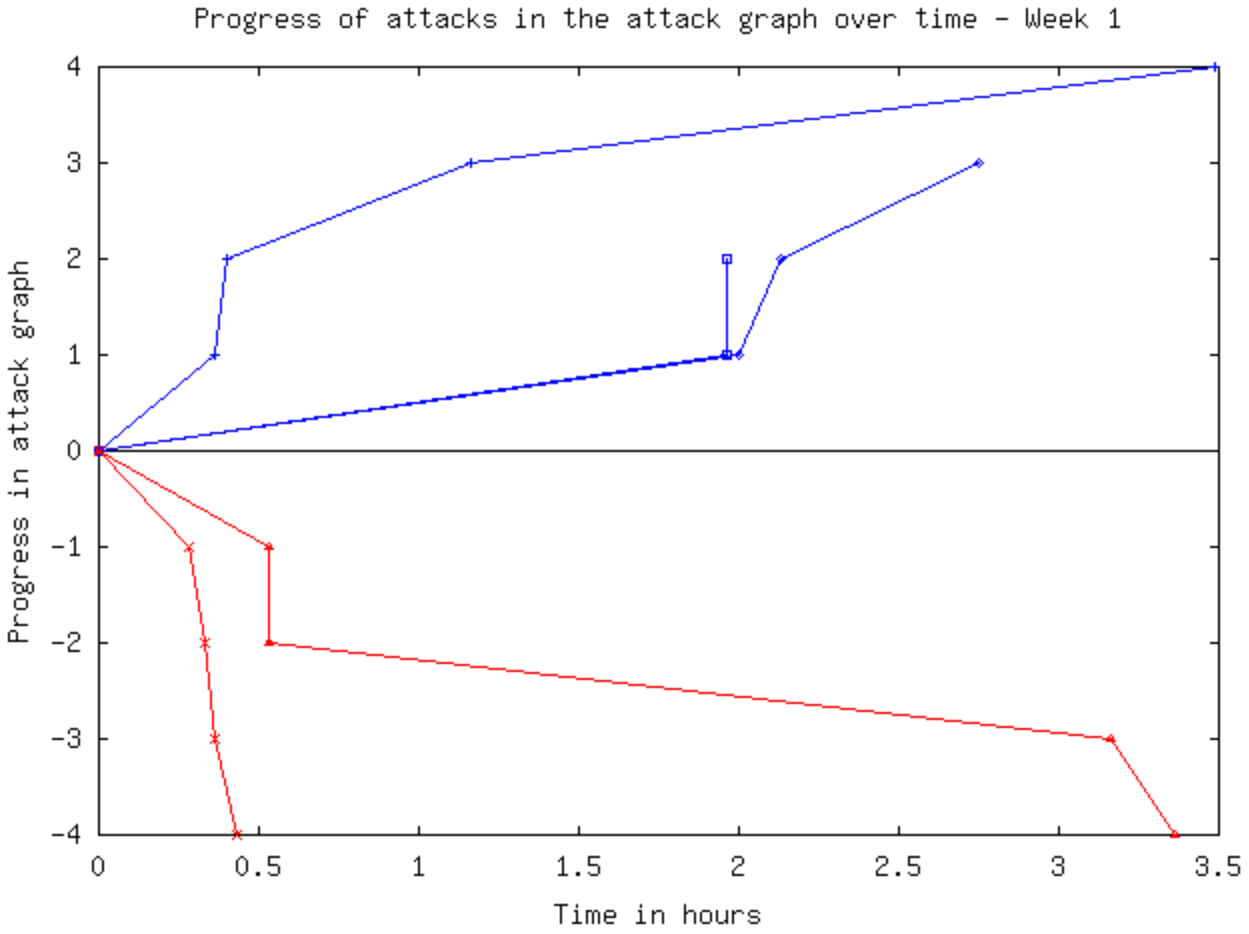
The first and perhaps most important thing to notice in the summary of results is that when deception is enabled, attackers never get as far toward the truth as they do when deception is disabled. In other words, deception works. Furthermore, it works very well. When deception is turned on, attackers almost uniformly go down the deception parts of the attack graphs rather than down the real parts of the attack graph. In cases other than blatant dazzlement, they are convinced that they are going down real paths for a substantial time. In some cases, attackers were so convinced that they had won when they were actually deceived, that they declared victory and walked away early. In some dazzlement cases, people got so frustrated that they gave up early. These results verify the previous anecdotal data from the Honeynet project [6] and Deception ToolKit [7].

The First Four Weeks of Experiments

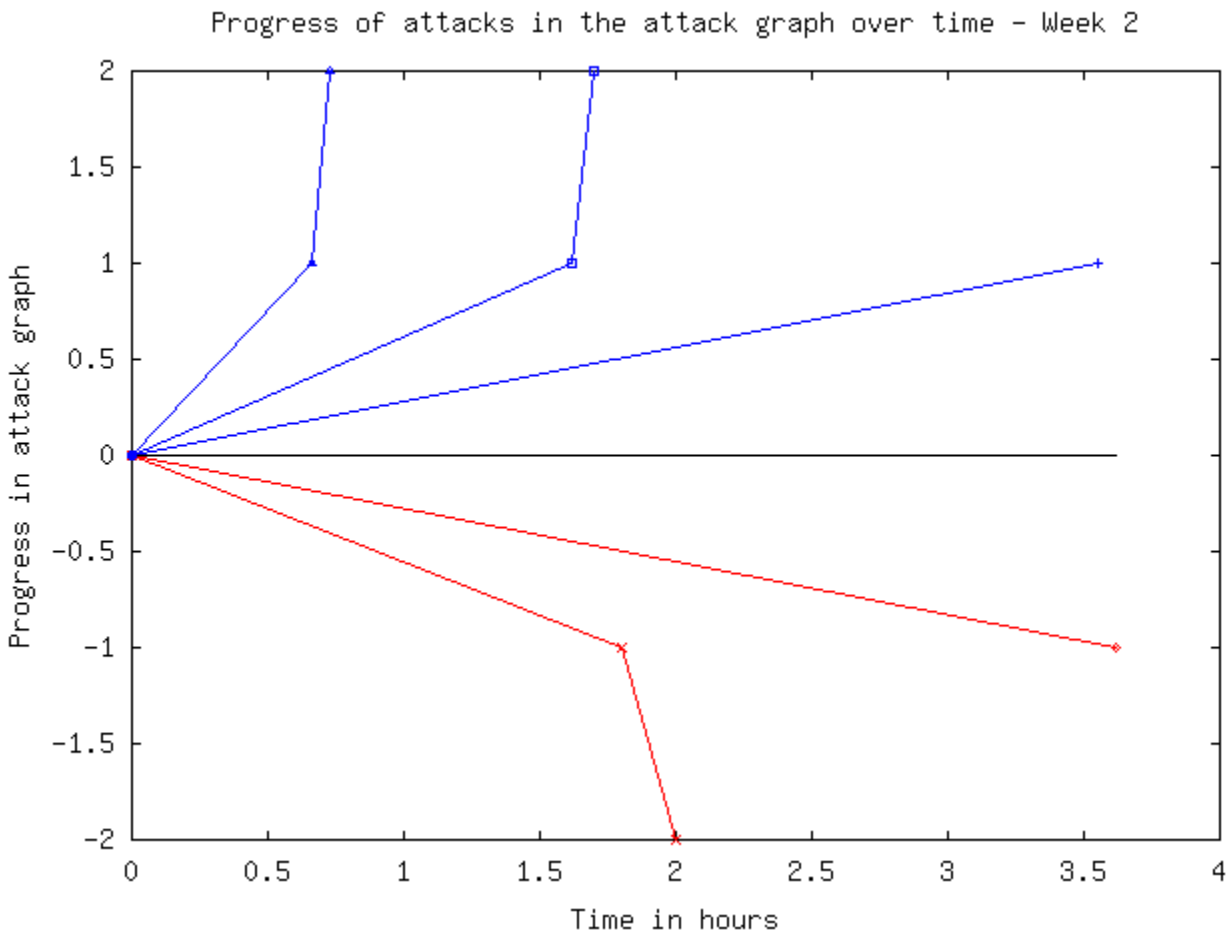
In the following plots, we examine each of the first four weeks of experiments, one week at a time. In the first three weeks, teams put earlier in the week were thought to be less able based on their known skill sets, no training was done for any teams, and a control group for each of non-deception (Tuesdays) and all-deception (Thursdays) were provided. This gave an advantage to the control

group with deception enabled because they were considered more highly skilled than the control group with deception disabled. Despite this advantage, deception was quite successful.

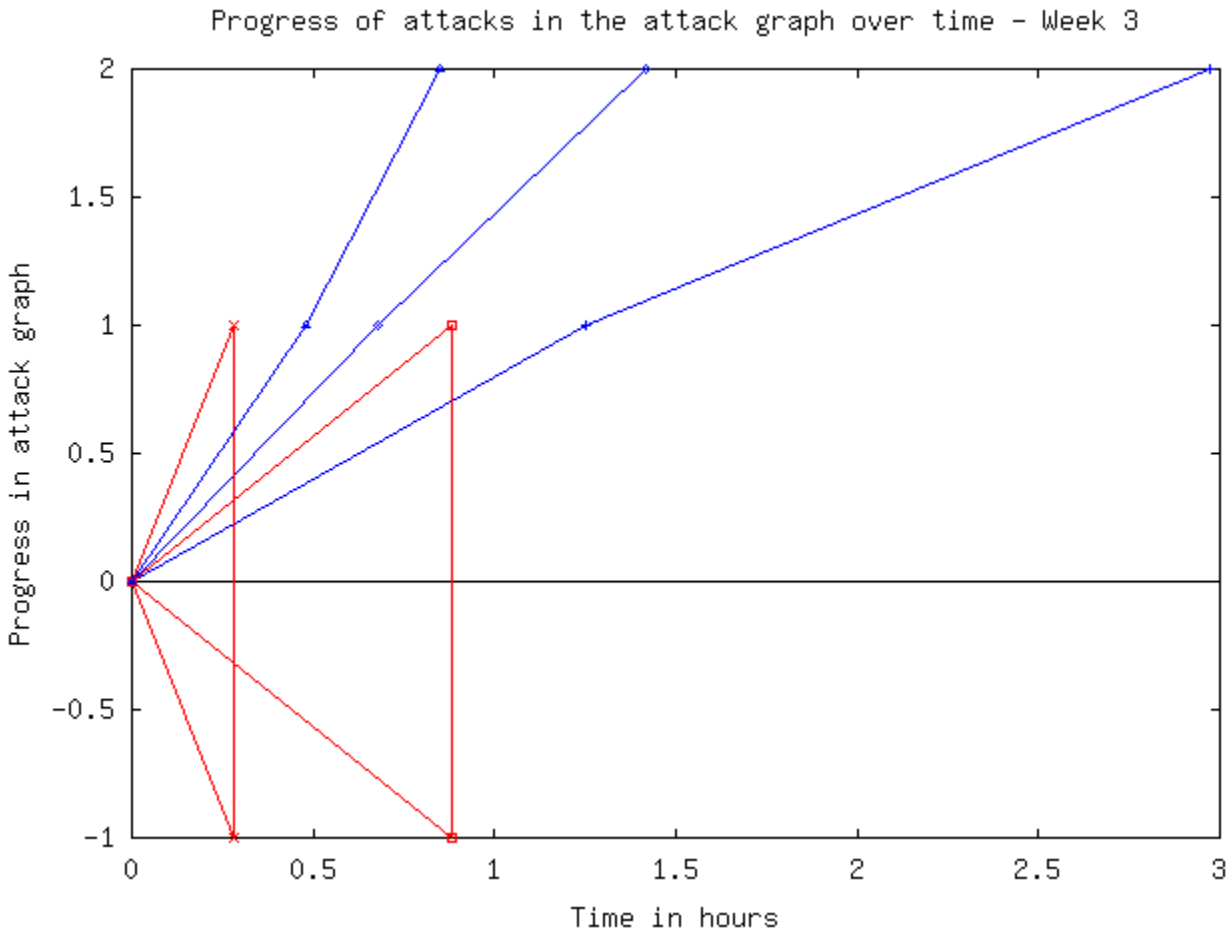
In the first week, we induced type 3 errors - errors of misdirection. This was done by introducing two identical systems, one obvious and available, and the other less obvious and harder to find. When deception was turned on, the obvious system was the deception while the inobvious system was to real object of the attackers' efforts. When deception was turned off, the only system to find was easily found and exploited. When deception was turned on, the obvious system was easily found and exploited. Since it appeared to the attackers that they had won, they gave up the effort and declared victory. The plot makes this very clear in that all of the non-deception attacks (in blue) succeeded, while all of the deception attacks (in red) failed but appeared to succeed.



In the second week, we induced type 2 errors - that is - errors of commission. This was done by providing a very large search space that was very thinly populated with real targets and, when deception was enabled, equally populated with easier to find deception systems. These experiments were designed to be open-ended in that it would be very hard for any team to get to the end in the time allotted. While no team got very far in these experiments, it is clear that the teams that ran without deception enabled made progress in the right direction while teams working with deception enabled made nearly the same amount of progress, but in the wrong direction. In other words, the rate of progress through the attack graph was similar but the deception was extremely effective.



In the third week, we induced type 1 errors - that is - errors of omission. This was done by providing an enormous quantity of false data - to the point where detecting possible targets was very easy but differentiating real targets from false ones was far harder. As we can clearly see, the effect of this deception was that attackers found both the deceptions and the real systems at about the same time but were unable to differentiate them and unable to make further progress. The attackers not undergoing deception were able to identify the real victim systems and make progress against them.



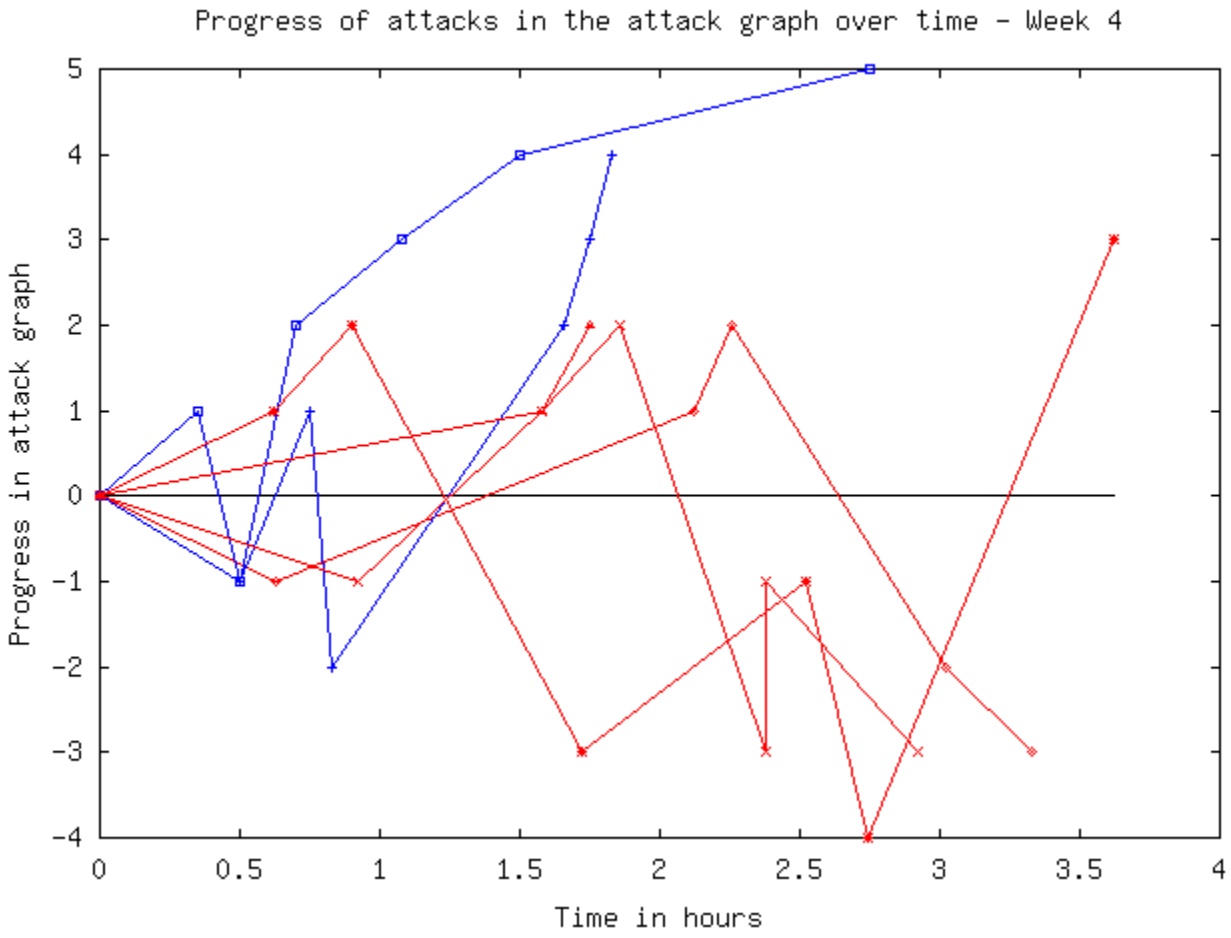
Week four represents the beginning of the second phase of the effort. For the first three weeks, the attackers under study were not told anything about deception or its use. While some of them guessed that deception was in use when they saw dazzlements, no information was provided from those performing the study. At the end of the third week, all groups were out briefed on the nature of the experiments and given full details about the nature of the research, the deceptions that were used, how they performed, and the technologies that were used. Starting in week 4, two changes were thus made. The first change was that the information provided about the situation included information about deceptions. The second change was that the next three weeks of attacks were performed against the same defenses. Thus teams could learn from previous efforts and attempt to avoid the mistakes of the previous week in the next week. Week four can thus be considered both in the context of the previous three weeks of experiments and in the context of the subsequent two weeks of experiments.

In week four, the system under attack was more complex than in previous weeks and for that reason we expect that total progress relative to the ultimate objective will not be as rapid. In fact the group that made the best progress resulted in a failure that could not be recovered from (the attackers caused an intermediary system required for success to fail). Interestingly, we see an increased variety of effects when the teams understand that deception is a possibility.

- One of the most startling effects is that teams suffer from self-deception. For example, the two teams that were not being deceived believed that they were being deceived at various times and acted on those self-deceptions. They performed additional experiments similar to those that someone being deceived would attempt and thus we noted these as deceptions in the

plot. They recovered fairly rapidly in comparison to teams actually being deceived, but this indicates that the mere threat of deception offers some protective value.

- A sixth team participated in this week's activities as well. This team consisted of the people who designed the experiments and included some of the people who had watched previous teams in these same exercises and who had almost complete knowledge of the manner in which the experiment was being undertaken. They had been previously briefed on the attack graphs including the deception paths and were extremely cautious in their approach. They included a senior intelligence officer (recently retired), two highly skilled systems administrators, a naval researcher, and a highly skilled security consultant who used to run intelligence operations for a state law enforcement agency. This group did not encounter any deceptions, and they made slow but steady progress toward their goal. Because of time limitations on the facility they had one hour less than the other teams and got further in the time alloted than the other two teams exposed to deceptions. The left very little in the way of footprints of their attacks, and while it is likely that they would have encountered deceptions in their next step, their experience and knowledge of the detailed attack graphs clearly benefited them. They did not, however, progress as far as the far less experienced teams that were not facing deceptions.
- Backtracking behavior was encountered among groups that were being deceived, and this resulted in revisiting parts of the attack graph that had previously been encountered and being (in one case) redeceived or (in the other case) deceived by a deception that had previously been avoided. The fist case is seen in the team that achieved -1 at 1 hour, -3 at 2.4 hours, and -1 again at 2.4 hours. The second case is seen where another team encounters -3 at 1.7 hours and then encountered -2 for the first time at 2.5 hours.
- The movement back and forth between real progress and false progress, between reality and deception, and between deception closer to and further from the starting point indicate that measuring progress toward the goal is far more difficult for the targets of the deception to assess because of the lack of clear and consistent feedback available by direct observation. The problem of counterdeception is clearly in play here and the need for some high assurance feedback for the attackers seems clear if progress is going to be made against such deceptive defenses.



Confounding Factors in the First Four Weeks

In our previous work [1] we identified a set of confounding factors associated with deception. Specifically, these are factors that affect movement between the three levels of cognition (low-level, mid-level, and high-level) identified in the previous cognitive model. The questionnaire that team members filled out after experiments and then filed out as a group combined with the observer's notes were intended to allow us to measure these factors. The data on confounding factors is analyzed to understand the relationship between these factors and performance.

The first summary indicates that the difference between results for all confounding factors when deception was on and off was fairly insignificant. The largest difference was 0.49 and this with a standard deviation of more than 1.1. In all cases the difference (shown in the last column) is less than 50% of one standard deviation. From this we can conclude that there was no clear relationship between the presence ofr absence deception and any of the identified confounding factors in these experiments. Strangely, time pressure and distraction were lower when deception was on than when it was off. While these differences are not very statistically significant, they appear to be present.

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time	Uncert	Dist	Tired	Hard	Int	Joy	Surp
Off	3.31	3.06	2.61	2.96	2.71	2.61	3.76	3.14	3.16	2.73	2.92	3.94	3.57	3.2	3.41
StdDev	1.08	1.34	1.22	0.87	0.89	1.15	1.15	1.1	1.05	1.22	1.17	0.94	1.02	0.87	0.79
On	3.15	2.83	2.63	2.73	2.73	2.6	3.68	2.65	3.43	2.35	2.75	3.9	3.58	3.23	3.28
StdDev	0.77	0.93	1.17	0.78	0.93	1.1	0.94	1.08	0.98	0.89	0.81	0.93	1.15	1.12	0.85
Diff	0.16	0.24	-0.01	0.23	-0.01	0.01	0.08	0.49	-0.26	0.38	0.17	0.04	0	-0.02	0.13

The Relationship Between Deception and Confounding Factors

If we examine the same data week by week, we find that, despite a relatively small total amount of data per category, there are only a few cases where the difference in mean between deception enabled and deception disabled exceeds one standard deviation.

- In week 1 and week 4, both original and new strategies reportedly worked better with deception off than with deception on, but this result did not hold across weeks 2 and 3 and the statistical significance is between the 'on' and 'off' standard deviations in both cases where it appears.
- The belief in success was higher in week 1 when deception was present than when deception was not present, but in week 4 deception caused reduced belief in success. The deception in week 1 was very effective in concealing the fact of deception, while in subsequent weeks deception was not only more obvious, but also believed and eventually known to be present. The importance of success fell off in week 4, but this will be discussed in more depth in the analysis of weeks 4-6 below.
- Time pressure was perceived as higher with deception on than off in week 1, but this did not hold for later weeks.
- Uncertainty was higher for weeks 1 and 3 with deception on, but not in weeks 2 and 4, and not to a very significant extent. Distraction was negatively correlated with deception in all four weeks, but not at a very significant level.
- Exhaustion was never an issue, but difficulty was believed to be lower in weeks 1 and 2 when deception was enabled, while it was higher in weeks 3 and 4 when deception was enabled. This may be related to the suspicion and eventual knowledge of the presence of deception that grew over time.
- Increased difficulty was somewhat correlated to increased interest and in week 3, interest was higher when deception was on, but generally interest was kept high throughout these four weeks of experiments.
- Enjoyment was negatively correlated to deception in all except the third week, where the increased interest and difficulty apparently drove the subjects to desire to meet the challenge.
- No significant difference in surprise correlated to deception was reported in any of the experiments.

We thus conclude that, for this sample, confounding factors had some significant correlations with type 1, type 2, and type 3 errors relative to the presence or absence of deception.

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time	Uncert	Dist	Tired	Hard	Int	Joy	Surp
Week 1 On	3	2.88	3.75	3.25	3.63	4	3.88	2.13	3.88	2.25	2.38	2.75	3	3.13	3.38
StdDev	0.76	0.64	0.89	0.71	0.92	1.07	0.99	1.36	0.64	0.46	0.92	1.04	1.07	0.83	0.92
Week 1 Off	3.6	3.1	2.7	3.2	2.9	2.8	4	3.2	3.4	2.3	2.9	3.4	3.2	3.4	3.8
StdDev	1.07	1.45	1.49	0.42	0.57	1.55	1.15	1.03	1.07	0.82	0.99	1.26	1.32	0.84	0.92
Week 1 Diff	-0.6	-0.23	1.05	0.05	0.73	1.2	-0.13	-1.08	0.48	-0.05	-0.53	-0.65	-0.2	-0.28	-0.43
Week 2 On	2.71	2.86	2.14	2.29	2.14	2.29	3.86	2.71	2.86	2.57	3	4	3	2.86	3.14
StdDev	0.95	0.9	1.07	0.95	0.9	0.95	1.07	0.95	1.57	1.27	0.82	0.58	0.82	1.07	0.9
Week 2 Off	2.79	2.57	2	2.5	2.21	1.93	3.57	3	3.21	2.93	3.36	4.5	3.29	2.93	3.14

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time	Uncert	Dist	Tired	Hard	Int	Joy	Surp
StdDev	0.89	1.28	0.96	0.76	0.89	0.83	1.02	1.24	1.05	1.07	0.93	0.52	1.07	0.83	0.77
Week 2 Diff	-0.07	0.29	0.14	-0.21	-0.07	0.36	0.29	-0.29	-0.36	-0.36	-0.36	-0.5	-0.29	-0.07	0
Week 3 On	3.7	3.4	2.8	3	3	2.5	3.9	3	3.3	2.3	2.9	4.3	4.4	4.1	3.4
StdDev	0.48	0.7	1.03	0	0	0.71	0.74	0.67	0.82	0.95	0.74	0.67	0.84	0.99	0.84
Week 3 Off	3.31	2.77	2.77	2.92	2.85	2.85	3.69	3.15	2.62	2.69	2.69	3.62	4	3.15	3.62
StdDev	1.25	1.36	1.36	1.26	1.21	1.14	1.55	1.21	1.26	1.75	1.44	0.87	0.71	1.07	0.65
Week 3 Diff	0.39	0.63	0.03	0.08	0.15	-0.35	0.21	-0.15	0.68	-0.39	0.21	0.68	0.4	0.95	-0.22
Week 4 On	3.07	2.4	2.13	2.47	2.33	2.07	3.33	2.67	3.53	2.33	2.73	4.2	3.6	2.87	3.2
StdDev	0.7	1.06	1.06	0.83	0.9	0.8	0.98	1.18	0.83	0.9	0.8	0.68	1.24	1.13	0.86
Week 4 Off	3.67	3.92	3.08	3.33	3	3	3.83	3.25	3.5	2.92	2.67	4.08	3.75	3.42	3.17
StdDev	0.98	1	0.9	0.49	0.43	0.85	0.83	0.97	0.52	1	1.23	0.79	0.87	0.67	0.72
Week 4 Diff	-0.6	-1.52	-0.95	-0.87	-0.67	-0.93	-0.5	-0.58	0.03	-0.58	0.07	0.12	-0.15	-0.55	0.03

The Relationship Between Deception and Confounding Factors Week by Week

The table below summarizes the results based only on the ratings of the confounding factors week by week. When deception was enabled, perceived success became worse with time, while when deception was disabled, perceived success became greater with time. Success was always considered important, but decreased slightly in import over time. Time pressure tended to increase over time for those under deception but not for those not facing deception. The lowest uncertainty was experienced with deception on, but generally did not correlate with the presence or absence of deception. Exhaustion was not correlated with these activities. All of the efforts were considered difficult to the participants with the exception of the first week which was very easy to complete, even if it was very hard to detect the deception. Interest and enjoyment were very high in the third week.

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time	Uncert	Dist	Tired	Hard	Int	Joy	Surp
Week 1 On	3	2.88	3.75	3.25	3.63	4	3.88	2.13	3.88	2.25	2.38	2.75	3	3.13	3.38
Week 1 Off	3.6	3.1	2.7	3.2	2.9	2.8	4	3.2	3.4	2.3	2.9	3.4	3.2	3.4	3.8
Week 2 On	2.71	2.86	2.14	2.29	2.14	2.29	3.86	2.71	2.86	2.57	3	4	3	2.86	3.14
Week 2 Off	2.79	2.57	2	2.5	2.21	1.93	3.57	3	3.21	2.93	3.36	4.5	3.29	2.93	3.14
Week 3 On	3.7	3.4	2.8	3	3	2.5	3.9	3	3.3	2.3	2.9	4.3	4.4	4.1	3.4
Week 3 Off	3.31	2.77	2.77	2.92	2.85	2.85	3.69	3.15	2.62	2.69	2.69	3.62	4	3.15	3.62
Week 4 On	3.07	2.4	2.13	2.47	2.33	2.07	3.33	2.67	3.53	2.33	2.73	4.2	3.6	2.87	3.2
Week 4 Off	3.67	3.92	3.08	3.33	3	3	3.83	3.25	3.5	2.92	2.67	4.08	3.75	3.42	3.17

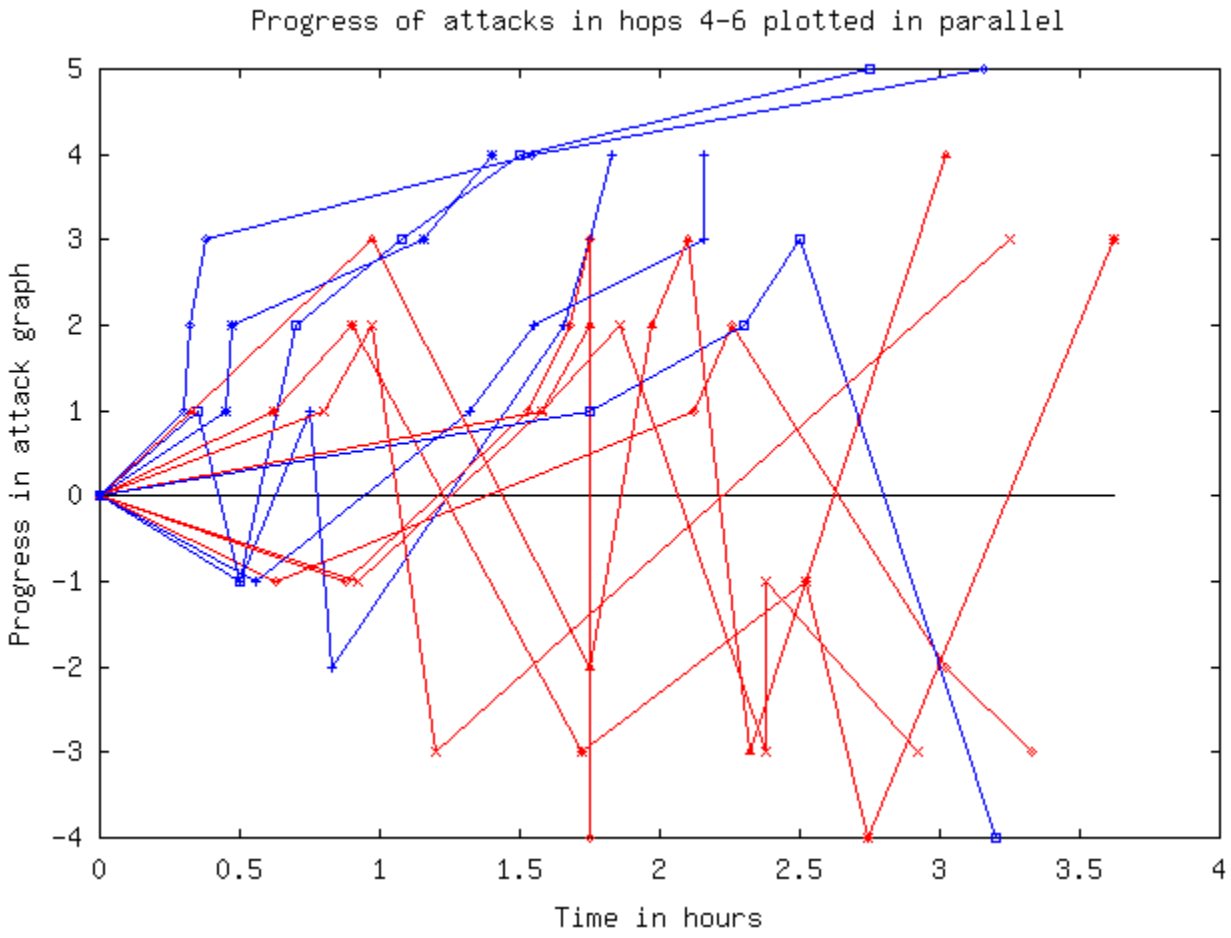
Magnitude of Confounding Factors Week by Week

More interesting results come in terms of difficulty, interest, enjoyment, and surprise. The first week was an extremely easy exercise designed to assure that all teams would believe they had achieved their objectives. The assessment of its ease would likely be more stark if they had the experiences in a different order, however, it is plain to see that it was easier by the results. Week 3 was considered quite difficult, most interesting, and most enjoyable by a significant amount. This particular exercise was pretty action packed in the sense that there were always things to see, things to try, and things going wrong. It was pretty interesting to watch as well. It was intentionally designed to induce errors of omission by providing massive quantities of information - more than could possibly be analyzed in the time allotted. Frustration ran high in a few instances, but clearly the participants enjoyed the effort, were engaged in the activity, and it appears that it drove them toward high-level cognitive activities.

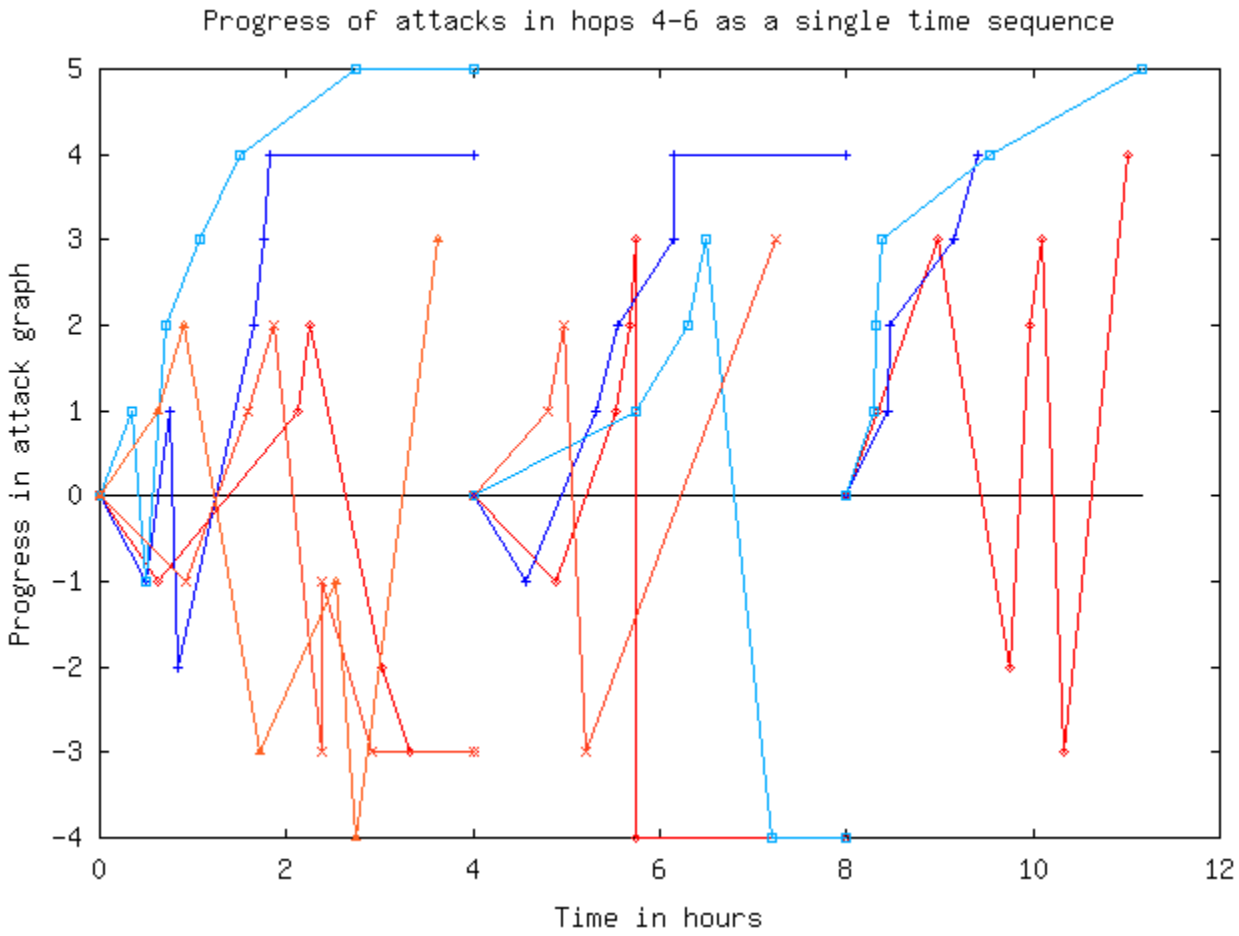
Experiments 4-6 Taken as a Group

There was a one week pause between week 4 and week 5 to allow teams to improve technologies in use and rethink the previous results. While the teams did not have their full time to spend on this effort for that period, all participants were also required to study attack scripts available over the Internet, were provided a series of training sessions on "red teaming", and were provided with classes on the use of command scripts for systems administration tasks. They indicated that this had a substantial improvement in terms of their skills. At the end of week 4, teams were provided with detailed reviews of the previous experiments including brief solutions on how they could go about defeating the defenses that were in place. This included specific details on how to defeat the systems in hop 4. In addition, insider supervisor access was provided to a system within the firewall for hop 4 so that attackers had detailed information comparable to that granted to an undetected attacker in an overrun situation. The attackers were also provided with full access to copies of all of the hardware and software systems used in the experiments, all network diagrams were accurate to the level of detail provided, and after the fifth week were provided with specific training and tools that would allow them to bypass the problems thay had the most problems with during the previous week.

When we plot hops 4-6 as if they were separate attacks we see that teams acting without deception working against them tend to get further in the real attack graph faster and that the effects of learning improve performance of attackers not exposed to deception. On the other hand, attackers exposed to deception do not appear to make progress deeply into the attack graph more quickly after being exposed to the situation repeatedly. They do remember some of their successes but they also appear to remember the deceptive information and continue down deception paths despite repeated exposure, additional training, increasingly detailed information about how the deceptions work, and increased training on attack tools.



When plotted as if the attack continued week after week with a 'reset' to level '0' at the end of each session we see that teams without deception learn to avoid self-deceptions and move through the attack graph more rapidly on subsequent tries, slowing as they reach the previous limit and moving forward slowly from there. Teams facing deception tended to give up and continued to go down deception paths even after more than 10 hours of effort. They repeatedly encountered the same deceptions and were unable to avoid them, apparently because they were unable to differentiate the deceptions from the real situation.



Slight Coloring Differences are used to Differentiate Groups

The Friday group decided to cease participation in the attack activity by the 5th week because it was perceived as a waste of time. In interviews with the test subjects, they indicated that they would prefer to spend their time on something more useful such as other projects that they worked on. This team (starting in week 4) consisted almost entirely of individuals who did not work every day of the week and thus had more limits on their time than other participants. It is not known whether this movement to other things represents an effect of deception, however, this team was operating with deception enabled during week 4 and made the most progress in their efforts of any team with deception enabled (they reached +3 at 3.5 hours after reaching -4 at 2.75 hours). Their forms indicated substantial frustration and high difficulty in week 4 as well as low enjoyment, all factors that we might predict would lead to resignation from the activity.

The Thursday team also decided to cease participation, but in their case this hapenned after the 5th week. They were also working against deceptions and they asserted that they were more interested in another student activity at that time. It is particularly noteworthy that this group decided to stop after reaching +3, just as the Friday group did. They first reached +3 at 3:15 in their second 4 hour session (or at 7:15 relative to the original start) and decided not to continue the next week.

The only remaining group working against deception (Monday) reached +4 only after 11 hours of participation and never reached +5. Both teams working without deception reached level 4 in the first 3 hours and reattained it before the Monday group to achieve it for the first time. The only group not undergoing deception to reach level -4 deceived itself by not ignoring its own packets in its analysis for a short period of time and recovered from this very quickly.

Deception clearly slowed the attacks, total progress against defenses is far worse when deception is present, and in this case, that attackers tend to abandon attacks in the face of deception while those not facing deception did not abandon the attacks.

Confounding Factors in Weeks Four to Six

We already mentioned that the Friday group abandoned the effort after confounding factors reached levels of 4/5 or above in their self-assessments. The following data shows the effects of deception on the confounding factors far more clearly. It is important to note that the number of samples became quite small at the end since only 3 out of the original 15 participants continued to participate (1 in 5). For the group not encountering deception, 8 out of 12 initial participants continued through the end of the sequence.

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time	Unc	Dist	Tired	Hard	Int	Joy	Surp
D On	3.1	2.79	2.24	2.59	2.31	2	3.38	2.24	3.55	2.79	3	4.38	3.34	2.86	3.28
SDOn	0.82	1.15	1.27	0.95	0.89	0.93	0.94	1.09	0.87	1.29	1	0.62	1.2	1.22	0.75
D Off	3.61	3.48	2.9	3.13	2.87	2.84	3.87	3.13	3.42	2.61	2.94	4.19	3.39	2.94	3.19
SDOff	0.93	0.82	0.76	0.63	0.48	0.91	0.86	1.06	0.77	0.84	1.13	0.76	1.07	0.94	0.81
OnOff	-0.51	-0.69	-0.66	-0.54	-0.56	-0.84	-0.49	-0.89	0.13	0.18	0.06	0.19	-0.04	-0.07	0.08

The Relationship Between Deception and Compounding Factors for Weeks 4-6

According to this data, the confounding factors related to the cognitive effects of deception are not strongly correlated to the presence of deception, but there is a correlation in some areas. For example, while surprise, enjoyment, interest, distraction, uncertainty, and difficulty were relatively uncorrelated to the presence of deception at this point, time pressure, desire for success, and planning indicators were negatively correlated with the presence of deception on levels at or near a standard deviation. This would seem to tend to indicate that an expectation of failure built up when deception was present, resulting in lowered expectations, less trust in planning and leadership, and, interestingly, less of a feeling of time pressure. As the desire and expectations of success were reduced, time apparently became less of an issue.

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time	Unc	Dist	Tired	Hard	Int	Joy	Surp
Week4-Off	3.67	3.92	3.08	3.33	3	3	3.83	3.25	3.5	2.92	2.67	4.08	3.75	3.42	3.17
StdDev	0.98	1	0.9	0.49	0.43	0.85	0.83	0.97	0.52	1	1.23	0.79	0.87	0.67	0.72
Week5-Off	3.5	3.36	2.93	2.93	2.79	2.64	4	2.93	3.5	2.86	3.29	4.36	3.14	2.71	2.86
StdDev	0.94	0.84	0.83	0.62	0.58	1.08	0.88	1.21	1.02	0.77	0.61	0.74	0.95	0.99	0.77
Week6-Off	3.88	3.38	2.88	3.25	2.88	2.88	3.88	3.13	3.25	2.13	2.75	4	3.25	2.75	3.63
StdDev	0.83	0.52	0.64	0.71	0.35	0.35	0.99	1.36	0.46	0.99	1.39	0.76	1.28	0.89	0.74
Week4-On	3.07	2.4	2.13	2.47	2.33	2.07	3.33	2.67	3.53	2.33	2.73	4.2	3.6	2.87	3.2
StdDev	0.7	1.06	1.06	0.83	0.9	0.8	0.98	1.18	0.83	0.9	0.8	0.68	1.24	1.13	0.86
Week5-On	2.91	3	1.91	2.64	2.09	1.64	3.55	1.73	3.64	3.64	3.55	4.64	3	2.82	3.45
StdDev	0.93	1.04	1.46	1.28	0.93	0.92	0.74	0.71	1.07	1.3	0.89	0.53	1.04	1.06	0.76
Week6-On	4	4	4	3	3	3	3	2	3.33	2	2.33	4.33	3.33	3	3
StdDev	1	1	1	0	0	1	1	1	0.58	1.73	1.15	0.58	1.53	1.73	0

Week-by-week Deception-differentiated Figures for Compounding Factors

Things get even more interesting as we examine the time effects of deception. Note that because a large portion of those undergoing deception opted to stop their efforts, the data values of those who did not participate are not present in the statistics when they are not participating. The removal of the

participants with the least interest and enjoyment are likely the reason there is not a large negative correlation of enjoyment with deception. In exit interviews those who left indicated that they were not enjoying the activity very much and that their interest was falling off in favor of their other work. Difficulty was perceived as very high for this effort by all parties, and particularly more difficult, distracting, and uncertain in the second week for those who subsequently left. As the perception of potential for success was reduced the teams also became less able to work together.

D	Team	SI	SW	NSI	NSW	Suc	ISuc	Time	Unc	Dist	Tired	Hard	Int	Joy	Surp
Week4-Off	3.67	3.92	3.08	3.33	3	3	3.83	3.25	3.5	2.92	2.67	4.08	3.75	3.42	3.17
Week5-Off	3.5	3.36	2.93	2.93	2.79	2.64	4	2.93	3.5	2.86	3.29	4.36	3.14	2.71	2.86
Week6-Off	3.88	3.38	2.88	3.25	2.88	2.88	3.88	3.13	3.25	2.13	2.75	4	3.25	2.75	3.63
Week4-On	3.07	2.4	2.13	2.47	2.33	2.07	3.33	2.67	3.53	2.33	2.73	4.2	3.6	2.87	3.2
Week5-On	2.91	3	1.91	2.64	2.09	1.64	3.55	1.73	3.64	3.64	3.55	4.64	3	2.82	3.45
Week6-On	4	4	4	3	3	3	3	2	3.33	2	2.33	4.33	3.33	3	3

Magnitude of Compounding Factors

The clearest change in characteristics happens at the break between week 5 and week 6. The uncertainty, distraction levels, tiredness, and difficulty factors all increase in the deception-on group from week 4 to week 5, and as an apparent result, in week 6 these participants leave and levels return to more moderate levels. Similarly, in this period, interest and enjoyment drop in the group of people who left, dragging down the average, which then increases after this group leaves. Similar events appear to correlate for the group that left at the end of week 4, however, the overall effect is less stark because there are more people in the sample space at this time.

Special Runs

In addition to the runs described above as series of 'special runs' were carried out both as demonstrations. These did not include control groups and were under unique circumstances. The results are illuminating and are thus provided here, but the reader is cautioned that they are each special cases and conclusions are thus more speculative for now. These results are not plotted with other results or collated in summary tables because they are not statistically meaningful. They do, however, represent examples of experiments that might be repeated under controlled conditions in the future so as to provide additional statistically valid results.

SR-1 was a slight variation on experiment 3 in which the increased rotation mechanism was not operating and the Trojan horses in the real victim system were very stark and obvious. It serves as a very good example of an insider team with local access because:

- It involved two more experienced attackers as participants, one with many years of experience in testing systems for vulnerabilities and another who runs higher quality attack research and development teams,
- It included team members who had previously done exercise 3, both with deception on and deception off, and a team member who had previously supervised experiment 3 as the observer, including doing the system configuration.
- False traffic was introduced using an additional method to better simulate a real environment of this sort.
- The experiment was done after all of the other experiments and included team members who had been through the entire sequence of previous experiments.
- Several team members had been briefed on deception and deception techniques as part of the project team,

- Training and tools were provided specifically on how to detect and bypass the sorts of deception in use, and
- Specific training on how to rapidly accomplish denial of service was given, including an example demonstrated in less than 15 seconds which served as the prototype for the actual denial of service attack used.
- In this case, the team was successful in 2R (in that they started to seek content), but doubted their results and never acted on them (thus 2D), and was successful in 3R but didn't fully believe it even as the experiment ended (thus 3D). All they convinced themselves of was that (1) they had found something that they did not believe to be what they were looking for and (2) that when they were running out of time, they denied services to themselves. This seems to indicate clearly that the knowledge of the possible presence of unknown deception has very beneficial defensive effects on potential attackers and it strongly supports the notion of publishing results on deception at this level of specificity.

Summary

Perhaps a more important result than this is that these deceptions were reasonably effective for a small period of time against non-expert attackers even though the quality of the dazzlement deceptions made them fairly easy to differentiate from live traffic. As we worked on improving this quality within fairly limited scope, we were surprised at how important the improvement in quality can be to the effectiveness of the deception.

Based on these results it appears that the network technology deception capabilities are very effective at what they do, but that in order to be far more convincing for a far longer time against more skilled attackers, it will be necessary to create improved content-oriented deceptions. The net objective of combined deceptions is that attackers spend more time going down deception paths rather than real paths, that the deception paths are increasingly indifferentiable to the attackers, and that the defenders can gain time, insight, data, and control over the attackers while reducing defensive costs and improving outcomes.

References

[1] Fred Cohen, Dave Lambert, Charles Preston, Nina Berry, Corbin Stewart, and Eric Thomas, "A Framework for Deception", Computers and Security, 2001 (submitted).

[2] F. Cohen, "A Mathematical Structure of Simple Defensive Network Deceptions", 1999, http://all.net (InfoSec Baseline Studies).

[3] Detailed research data is not available for reasons of participant confidentiality.

[4] Red Teaming Questionnaire Form

[5] Standard Red Teaming Pre-Briefing

[6] The HoneyNet Project web site (www.honeynet.org).

[7] Fred Cohen, "Deception Toolkit", March, 1998