VOLUME 2 NUMBER 2

JOURNAL OF
CYBER SECURITY &
INFORMATION SYSTEMS

Games People Play

Behavior and Security

CSIAC

Toward Realistic Modeling Criteria of Games in Internet Security

By Jonathan M. Spring

There have been various attempts to apply game theory to various aspects of security situations. This paper is particularly interested in security as relates to computers and the Internet. While there have been varying levels of success in describing different aspects of security in game-theoretic terms, there has been little success in describing the problem on a large scale that would be appropriate for making decisions about enterprise or Internet security policy decisions. This report attempts to provide such a description.

We propose that there are three types of players in the game: the computer user, the malicious actor, and the security architect. This paper is not about how to "win" the game of Internet security or a prescription of the clever strategy — as game theorists make clear, "the search for effective decisions is not a central problem of game theory" [29]. The aim of this paper is two-fold, one for theorists and one for practitioners. For game theorists, this paper provides a more accurate description of the actual dynamics of security-related interactions on the Internet. For practitioners, we will provide a framework to clarify existing motivations and intuitions about the current situation and why it is, or is not, working. Hopefully this perspective on the dynamics of the situation will enable more effective decisions and guide the search for clever solutions using other fields of study.

This paper does not focus on building mathematical tools for analysis. We focus on the description of the game. The three players — user, rogue, and architect — all have competing interests. The main interactions are thus: (1) The user and architect negotiate a suitable system configuration which includes trade-offs between productivity (of the user), security (architect's goal), and cost; this is a non-zero sum game. This occurs on a much slower time scale than the other two interactions. (2) The rogues attempt to steal resources from the user; this feature is also not a zero-sum game, and so presents some interesting challenges. (3) The third interaction is between the architects and the rogues. Although these two parties are defined as diametrically opposed, their interaction is also not zero-sum.

With these interactions laid out, we make the following important observation about the game itself: the user can ignore, or even be complicit with, the rogue without immediate loss. This fact makes it harder to convince the user to work with the architect to improve security. There are other interesting points to consider related to the game: (1) The game is modeled with three players, and we assert that at least this many players is necessary to maintain fidelity with the real Internet; (2) perfect security cannot be promised, even in principle, because the features of the game are such that there is no guaranteed method to compute a globally-optimal strategy (three player game, the fact that it is non-zero-sum, and the fact that there is imperfect information).

1 Introduction

Game theory was founded as a sub-discipline of mathematics in the mid-20th century. It is a description of how rational decision makers compete. However, this paper is not about how to "win" the game of Internet security or a prescription of the clever strategy — as game theorists make clear, "the search for effective decisions is not a central problem of game theory" [29]. What game theory can illuminate is how an interaction proceeds, certain rules about the outcome given the inputs, and to help an analyst clarify a situation by reducing a complex situation to a more compact description.

For the purposes of this paper, we will assume the payoffs to the players are already defined. How to do this is non-obvious. However, a process such as the model described in

[35] provides a plausible method for arriving at the payoffs, measured in monetary resources lost or gained.

Game theory assumes we have rational decision makers. Kahneman's psychological work, and the resulting behavioral economics literature, demonstrate that people are not purely rational. This has important ramifications for actually selecting policies that will be effective, however from our abstract point of view it just means we might have to adjust our payoff values to account for the fact that people may value something more or less than is rational. As such, we will leave this issue aside for now.

When describing the game, we will describe the payoff matrices to the extent possible — which values are positive or negative, their relative magnitudes, etc. However, our goal is not to formulate games to the level of detail that analytic or numeric solutions are possible. There is still much work to be done before that can be achieved. The goal of this paper is to provide the shape of a game as it relates to information security on the Internet.

2 Related work

Game theory was kicked off in 1944 as a robust field by [37] and saw application to such national security issues as nuclear deterrence and mutually assured destruction. The essential problems of bargaining and non-cooperative games were laid out by John Nash in the early 1950s [27, 26]. Founded as a branch of mathematics, after the theory acquired conceptual foundations (see [30, 29] for a summary), notions from game theory spread to a number of fields, notably economics (for example [31]). Some game theorists have also taken influence from other fields, such as evolution and dynamical systems [15]. Some game theory texts are broad, mathematical treatments such as [28]. Useful for the work described in this paper are treatments of non-cooperative games and games of incomplete information, which is included in some of the above but focused in some texts such as [14, 25].

There have been previous efforts to extend game theory into the field of information security; [34] summarize and categorize the efforts. Game-theoretic models have been proposed for both organization-scale [7] and single-wireless-node-scale [40] information security games; both as single-play [36] and repeated games [20]. As economics intersects game theory it also intersects information security; for a summary of the extensive work on the economics of information security, see [5].

We heuristically derive our model from case studies and empirical reporting of information-security relevant behavior on the Internet. There are several organizations that report on various aspects of cyber-crime and human behavior, in varying levels of detail, such as [4, 33, 2, 23, 13, 19, 24, 9, 8, 1]. These sources do not generally attempt to derive a general model from the information observed. There is some work in cyber-crime and risk dynamics such as [23, 35] that model criminal behavior, which inform our game theoretic modeling directly.

It seems that all existing applications of game theory to information security force the game to be a two-player game. Some study population dynamics of users and adversaries [39], which has richer descriptive power, but these retain still only two types of players. These efforts do not seem satisfactory in describing the Internet-scale phenomenon of information security, as reported by the economics, cyber-crime and dynamics literatures. We assert that a primary reason for this shortcoming is that the game cannot be described with fewer than three players.

3 Theory

The following subsections describe a more adequate treatment of the modern Internet security landscape. First, we describe the players; secondly, how they interact informally; finally, a more formal definition of the interaction.

3.1 The players

We shall define three classes of players. Granted, these classes may be subdivided for certain purposes, but we shall treat them as the essential units for our purpose of providing an accurate and useful model of the security interactions on the Internet. A single person or machine may change roles during its lifetime, and the ability to do so presents practical challenges, however we shall treat the three classes of players as describing mutually exclusive and exhaustive roles. The first step is to describe these players, their goals, and their capabilities.

User is an agent who utilizes a computer system. By definition, they have not designed the system they are using.[1] The user may have access to a limited number of configuration options provided by the architect of the system. The main goal of the user is to produce some product of value, using the computer

1 An agent may both use one system and be the architect of another; most software developers fit this description. However the roles of user and architect qua roles do not overlap.

system as a tool to that end. Possible products span the range of human ingenuity. An important consequence is that the Internet as we are analyzing it is not a closed system, it is a tool of the larger human economy. This is a factor in the assessment that games involving users are not zero-sum.

Architect is the agent that has designed a computer system or the policy under which the system operates. This may be operating system developers, enterprise security policy designers, or the IETF; there is a wide range of systems and they all have architects. Architects can also be identified with the owner and administrative operator of a system, especially in the case of enterprise organizations. The architect is who selects and enforces security policies.

Architects, as a group, are the hardest to unify as one label. Members of this group are highly specialized and fractal. Since no organization builds all of its own software, every architect is also the user of other systems. However, the essential element is not what role a particular person has. The key fact is that every system has an architect or architects that have designed it. The Internet is not a natural phenomenon, and so while it is bound by some physical laws the key feature is that every system that operates on the Internet has an architect who made decisions about that system, its capabilities, and so on. In the general case, the architect's goal is for their system to be used by users. A part of this goal is making it secure enough to be used, however it would be naive to say that an architect's primary goal was a secure system. If this were the only goal, the systems could all be turned off and encased in concrete to accomplish the goal. To specify what it means "to be usable" the architect specifies aims in reference to what users need to accomplish user's goals.

Rogue is the attacker. The definition of an attack can be disputed, but we shall mean attack as defined by Howard and Longstaff: "a series of steps taken by an attacker to achieve an unauthorized result; ... among these steps is an action directed at a target (an event), as well as the use of some tool to exploit a vulnerability. Second, an attack is intended to achieve an unauthorized result as viewed from the perspective of the owner or administrator of the system involved" [33]. Considering a system as large as the Internet, and given global political disagreements, it would be naive to think that we could agree on one rogue or set of rogues. An entity that is a user according to one point of view may be a rogue according to another point of view, and we may not be able to say that either point of view is correct. However, each user will experience a rogue that perpetrates attacks. The scope or goal of these attacks may vary; money, fame, chaos, or national interest may all be motivators for different rogues in varying degrees.

We assert not only that these are three players in a game describing Internet security, but that these are the only three types of players. Further, that all three must be modeled if a description of Internet security is to be accurate. For discussion of modeling more than three players at once, see Subsection 3.2.4.

3.1.1 Realizations

There will be multiple realizations of this game occurring in the world simultaneously. There are more than three agents on the Internet at any given time, so the above is a simplification. From different points of view different agents will be considered to be in different rolls, whether user, architect, or rogue. The fact that agents can change roles certainly can have real-world impacts. For example, the NSA's involvement in the design and architectural review of DES can be seen from many points of view [18]. The historical claim was that NSA may have negotiated down the key size because it wanted to be able to attack the protocol more easily in its role as rogue. However, it was permitted at the negotiating table in the first place because it was going to be a legitimate user of the protocol as well.

That one organization may have competing goals does not break the user-architect-rogue model. Various realizations of the model in the real world will alter the agents playing the game, their rolls, and their payoffs. However, the changes do not affect the general description of the game. This is one benefit of describing the game at the proposed level of abstraction. Although agents can serve multiple roles simultaneously in the real world, to simplify modeling at this early stage we put that aside and focus on the goals of the agents in each of the three rolls and the essential features of the interactions between a user, architect, and rogue.

3.2 Features of the Game

It is not possible to separate the three players from each other. The observation that there are three distinct, essential roles in the game is a vital observation. The other attempts (see Section 2) at bringing game theory into security have focused solely on two-player games. We believe this to be a primary factor in why these attempts have had unsatisfactory applicability to actual security decisions.

A three-player game results in a three-dimensional strategy space. Such a space can be difficult to conceptualize. In order to introduce the dynamics more gradually, we describe the essential features of each two-way interaction separately before combining this into the holistic problem.

One feature common to all three interactions is that they are not games of perfect information. The user does not know everything the architect has done, and vice versa, and likewise with the rogue. This is true in principle for the rogue, but it is also true in practice for the user-architect interaction. Security best practice such as least privilege and least access, legal standards of privacy, technical limitations on data processing, and the use of closed-source programs all make imperfect information a practical reality that is ingrained in the day-to-day use of the Internet. In fact, one plausible negotiating point is how much visibility the architect has into what the user is doing, and so on. This means the game is not guaranteed to have a globally-optimal strategy, as only games of perfect information are guaranteed to have one. The gist of the interactions of the players is summarized in Figure 1.

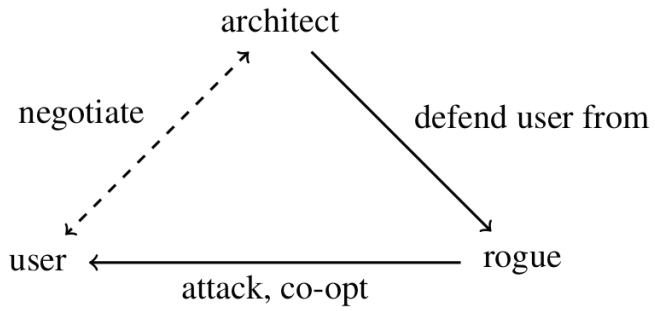


Illustration 1: Representation of the three distinct types of players in the proposed game. Dashed line indicates a game that is repeated at a slower pace. Labels on the edges describe the intent of the interaction. Although the arrows indicate that, for example, "architects defend user from rogues", this interaction is not independent of the others; all three players play simultaneously.

3.2.1 user-architect

The user and the architect are negotiating features of the system being used. Either side may be advocating for adding, removing, or modifying features of the system. Security requirements and rules can easily be viewed in this space. The simplicity of the statement hides a degree of difficulty in game-theoretic terms, however negotiation games have been reasonably well studied [6].

This part of the game is non-zero-sum; the user and architect can clearly come to agreements which are better for both of them. While the interaction is not antagonistic, it is not truly cooperative, either. The user and architect have different, ostensibly unrelated goals. So we should expect the user and architect to cooperate only insofar as it is mutually beneficial based on the payoffs provided to each.

One common aspect of both game theory and economics is the idea of discounting future payoffs in a repeated scenario. Colloquially, this is captured by "a bird in the hand is worth two in the bush." Rational decision-makers will value an equal payoff now rather than later if they have an expectation the game will end or change before that future payoff [31]. This is precisely the scenario we are building here, as the payoffs will be renegotiated at future points. Discounting is a reasonably well-studied feature in game theory. One important aspect that is practically important is that different entities can have different discounting rates; that is, entities are not equally patient [31]. To model the dynamics, a valuation of the initial capital of the parties is also necessary, which would have to take into account physical and information assets.

Not only does this user-architect interaction involve variable discounting rates by the parties, but the payoffs going forwards are also a function to some extent of investments made by players in the past. For example, if the user wants a capability in a system that does not exist, the architect will have to build that capability over a period of time. This requires resource investment before the benefits of the capabilities can be realized. Game theorist have studied games in which the players' past actions affect future payoffs, especially in the context of financial investment [16]. Although multiple investments could be modeled, in Section 3.3 the investment that is modeled is the infrastructure controlled by the rogue, which the user and architect have an interest in minimizing and the rogue wants to maximize.

One interesting characteristic of the user-architect interaction is that what the two parties are negotiating boils down to the payoffs for the parties in the user-rogue and architect-rogue interactions. Realistically, every several months system configurations could be renegotiated, however the other two interactions occur on second-to-second time scales. Conceptually, the user and architect negotiate payoffs in a repeated game every so many plays of the game. How often renegotiation happens would also probably be a feature of the negotiation. Whether this can be modeled as a situation in which the user and architect usually only have the option to "change nothing" at most stages of the game is not known. In principle, this could

be done without loss of generality or specificity. In practice that approach seems unrealistic — the user and architect do not check in every few seconds to confirm "change nothing" — but it may be a feasible model.

However it is modeled, the players are assumed to be bound by the terms negotiated for some number of repetitions of the faster games. Another possible option for modeling the problem could be borrowed from multiscale mathematics [38]. Unfortunately, we are unaware of any applications of multiscale mathematics to game theory at this time.

3.2.2 user-rogue

Conceptually, the interaction between the user and the rogue is one in which the rogue is attempting to steal the user's resources. Since this is theft, it would appear on the face of it to be a zero-sum game. However, we do not believe this to be the case. The rogues are not necessarily stealing purely rivalrous goods. If the resources stolen are non-rivalrous, then the user is not inconvenienced by the rogue's usage, and so the game is non-zero-sum. Money is rivalrous, but money is not the only resource the rogues steal. Rogues can also steal computer resources or information.

For an example of rivalrous and non-rivalrous goods, consider a sweater. If it is cold out, I like to wear a sweater. If you steal my sweater, I cannot wear it and I will be cold. Sweaters are a rivalrous good. Stealing my sweater would be a zero-sum game, because one's loss is precisely the other's gain. Now consider Pythagoras's theorem concerning the lengths of sides of a triangle. If a teacher knows it, and teaches the students that $a^2 + b^2 = c^2$ the teacher is not excluded from using that information. It is not as if the teacher gave out 20 sweaters. The usage of the theorem does not prevent others from also using it. Theorems, and information items generally, are non-rivalrous.

Internet access and computer processing cycles are not precisely the same as information in this regard, but they are more alike to non-rivalrous goods than they are like sweaters. If a user is only consuming 10% of available Internet bandwidth because, perhaps, they are asleep or out of the house most of the time, then a rogue with control of the computer can use the rest of the bandwidth without inconveniencing the user. Likewise with processor cycles and disk storage space. Precisely how the rogue must act in order to achieve this goal may require some technical cleverness, however here we are interested in specifying the nature of the game, not clever ways to attack or protect systems.

Information (and computer resources) can be given a monetary value. Information is often given monetary value in intellectual property rules and debates, for example. Yet the same information may easily have different value to different parties. Thus while we may reasonably expect to value the resources in the game we are describing, the game will be non-zero sum not just because the goods in question are non-rivalrous but also because the different players value the resources differently. For example, even if Eve gains something that Alice loses, if Eve considers it to be worth 1 unit, yet Alice valued having it at 2 units, the transfer is non-zero sum.

The payoffs for this repeated game accrue on a relatively short time scale compared to the user-architect interaction, as noted above. Also similar to as noted above, the rogues likely have different discounting rates than either the users or the architects for the repeated aspect of the game. Given the illicit nature of the rogues' activity, it is plausible that the rogues are the least patient.

Prior actions will have an effect on future payoffs in this interaction. As the rogue compromises more user resources, the rogues can use those resources to compromise further user machines. These invested resources also play in to the architect-rogue interaction because rogues can use these compromised resources to evade the architects. Therefore, actions taken in this plane of the game directly affect others, just as in the user-architect interaction.

3.2.3 architect-rogue

This plane of the game describes the interaction between those who design and own the systems and those who are attempting to subvert those systems. It is the more difficult plane to characterize intuitively. Since all of this occurs on computer technology, the rogue can directly attack the systems the architect is using to protect the users. However, in order to maintain clarity, the rogue is not attacking the architect directly; they are attempting to subvert the protections that the architect has in place to protect the user. This includes aspects such as email filtering, anti-virus signatures or other host-based protection systems, firewall rules, intrusion detection system (IDS) signatures, etc. The architect can generally reconfigure these rules at machine speed, and the rogue can likewise employ countermeasures quickly, and so this interaction's timescale is approximately the same order of magnitude as the user-rogue interaction.

This game is clearly non-cooperative and is not a game of perfect information. Both parties are intentionally obscuring

their methods from the other. Even though the two players are directly competing to block or use resources, we posit the game is non-zero-sum. The architect does not directly lose resources if the rogue can successfully steal resources from the user, nor does the architect directly gain resources if the rogue cannot use them. In fact, the architect will generally have to expend resources to block the rogue.

Just as the rogue can build up resources in the user-rogue interaction as a kind of investment in future payoffs, the architect-rogue interaction can effect that investment. In this case, the effect is that the architect can reduce or block the resources available to the rogue. This is precisely the same set of resources that the rogue is building up in the interaction with the user; it seems reasonable that these invested resources could be modeled together, as described in Subsection 3.3.

3.2.4 More than three players

To this point, we have taken for granted the simplifying assumption that there are only three agents — a user, an architect, and a rogue. In order to model more agents, the modeler could consider each of these groups a coalition of agents. Then there is a coalition of users, a coalition of architects, and a coalition of rogues. Individual agents may switch coalitions if it is in their interest to do so and they are permitted to do so by the other members of the coalitions.

This coalition model would be able to account for some of the complexities of the modern Internet. No architect would wittingly negotiate their system settings with an adversary. However, if the users are a coalition of users, an agent may be able to enter the coalition of users and corrupt the negotiation process and then enter the coalition describing rogues to attack the system.

Admitting coalitions and arbitrarily many agents complicates models, and these added complexities would cloud the initial formulation we are pursuing here. For the time being, we will resume the simplifying assumption of three agents in order to describe the game requirements satisfactorily at this stage of development. Generalization to coalitions of users, architects, and rogues should be possible in future work as necessary to improve the expressiveness of the model.

3.3 Formalizing the game

The description provided in Subsection 3.2 now permits us to describe the game more formally. Conventions for mathematical representations are drawn from [29], which also contains an accessible explanation of the symbols.

3.3.1 Initial definition

First, we can posit that there will be payoffs accrued to each of the three players. We may represent these as *a* for architect, *u* for user, and *r* for rogue.

Secondly, there are some general state variables that will be held across the game. Namely, the vector $\vec{v}$ and the matrix S:

$\vec{v}$: the infrastructure available to the rogue

S : the payoff matrix that will be used for the fast-scale games, based on system configurations the user and architect negotiate

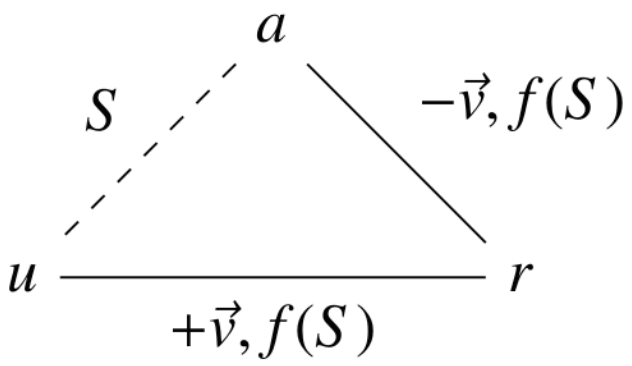


Illustration 2: Representation of the three distinct types of players in the proposed game. Dashed line indicates a game that is repeated at a slower pace. Labels on the edges denote values that are involved in each interaction; either the values are adjusted by the result of the game or the result of the game is a function of that value. The sign of the effect is noted if it is constant.

The rogue's infrastructure is really two things, hardware and software. Hardware are physical computing resources. This includes various incommensurate computing features, such as network bandwidth, processing power, stable disk storage, and volatile memory (e.g., RAM). In general, in both benign and malicious cases software is used to control this hardware. However, in considering rogue malicious infrastructure, software means software which has been developed in order to attack computers and wrest control from their rightful owners. In this sense, software does not mean how many machines each copy of software is deployed upon, but rather the total engineering effort the rogue has at their disposal.

One might expect a set of two vectors, one for hardware and one for software; this is not what is proposed. Both elements of malicious infrastructure are captured by $\vec{v}$ in different ways.

Modeling hardware is relatively straightforward. Hardware is the set of all machines in question, and a continuous value from 0 to 1 as to how much of the machine's operations the rogue controls. Being able to unplug the machine counts as control, so they do not likely have full control over many machines. We can thus define $\vec{v}$ more rigorously as:

(1) $\vec{v}$: = $[v_1, v_2, ..., v_q]$; $0 \leq v_\kappa \leq 1$

Where q is the number of computers in use. The operations of a specific computer κ are controlled to the extent v_κ by the rogue, where 0 is not at all and 1 is completely. Thus v_κ is a value for the percentage of the computer's resources that are controlled by the rogue.

How to model malicious software infrastructure is less clear. Software is a more complex set of capabilities the rogue has at their disposal. The architect is likewise constantly developing new software, patching vulnerabilities, and so on. Some architect activities, including detecting malicious software and patching vulnerabilities, do reduce the software infrastructure available to the rogue because these activities make certain malicious software ineffective against the rogue's targets.

One might imagine defining a matrix of the software available to the rogue and its effectiveness in targeting and maintaining control of each computer κ which has a representation in $\vec{v}$, perhaps similar to methods of modeling physical combat such as [22, 11]. We choose to simplify our model somewhat, and will consider these software interactions as part of what determines the payoff matrices and the changes to $\vec{v}$, but will not model them explicitly at this time. Modeling software infrastructure would be important future work in quantifying the game with realistic numbers from observation, but it is not necessary to understand the shape of the game as is described at the present level of abstraction.

An important feature is that the number of effective players in the game cannot be reduced by coalitions between the players. Although it is possible for the players to jointly improve their payoffs, because the game is non-zero-sum, the players cannot coordinate their actions in any coherent manner.

The rogue cannot cooperate with the user because they do not have an effective means to communicate or enforce agreements. The rogue is, by definition, achieving an unauthorized result. If the user were in a position to authorize the rogue's actions and come to an agreement, it would not be an attack scenario.

The architect and the rogue cannot be collapsed, although in the worst case for the user the architect and the rogue both are trying to undermine the user. Practically, if the architect designs a weak system the user may have few or no options that do not permit the rogue to perform a successful attack. However, the architect and the rogue have very different relationships with the user. The user and architect negotiate the payoff matrix, representing features of available software. The user will not wittingly negotiate with the adversary. And in most cases, the rogue also attacks the architect in order to bypass security controls, and so the same logic applies as to the user-rogue interaction. The architect cannot cooperate with the rogue because, by definition, the rogue is achieving an unauthorized result.

The user and the architect are practically prevented from forming a coalition because they cannot adequately communicate, share information, or enforce binding commitments. Thus, while the two players might wish to cooperate, and may be able to signal their intent to improve their situation [31], neither player is bound to cooperate. This situation seems similar to the classic "Battle of the Sexes" game [30]. The two players' interests do not align, but failure to agree — even when unilaterally choosing their preferred option — is worse than even the less preferred choice when agreeing.

3.3.2 Payoffs

We will not consider payoffs to be transferable and conservative, although arguments for doing so are plausible. A transferable payoff is one such that one player could transfer it, partly or wholly, to another player readily and without loss of value. Some of the elements of the payoff, such as time or computer resources, are not transferable. Some, such as money, are. If payoffs are transferable, bargaining becomes easier to analyze [29]. We do not believe this is accurate in our case. Since the players are not forming coalitions, non-transferable payments are less important, since most transfers are modeled as bargaining within coalitions or to entice a player to join a coalition. Even though money is a payoff, when involved in cyber-crime there is usually no practical way for the rogue to make payments to the user or the architect. For our purposes, we will thus simplify the payoffs and consider the payoff space S to be non-transferable, and thus exhaustive of all payoff options.

Each entry in S is a four-tuple, or an element with four distinct data elements. This includes the payoffs to each of the three players and the rogue's control over infrastructure vector, $\vec{v}$. Thus, S may be defined as:

(2) $s_{ijk} \in S; 0 \leq i \leq l; 0 \leq j \leq m; 0 \leq k \leq n$

(3) s_{ijk}: = $(p_a, p_r, p_u, \vec{v})$, *where* $p_a, p_r, p_u \in [0, 1]$

The values in each element s_{ijk} are the normalized real-number payoffs to the architect, rogue, and user, respectively, as well as the changes to vector $\vec{v}$ that is the outcome of that choice. The indexes i, j, and k are finite — at the present time there is no reason to believe the game has an infinite solution space. These values are the total number of strategies available to each of the architect, rogue, and user, respectively, at each step of the game.

The elements i, j, and k each indicate a strategy chosen by one of the players. Each time the game is played, each of the players receive the payoff at s_{ijk} and $\vec{v}$ takes on the value at s_{ijk}. The element s_{ijk} is selected out of S each time the game is played according to the strategies the players choose. Thus, the architect chooses the value of i, the rogue chooses the value of j, and the user chooses k. Strategies are chosen simultaneously. Each player chooses the value that will maximize their payoffs, however since s_{ijk} also affects $\vec{v}$ this consideration is more complex than usual. Players will consider investment and discounting when choosing their maximum payoff.

The payoffs themselves are represented as a function of $\vec{v}$. In this way, the payoffs can change in between renegotiations. More properly, the payoffs in each play of the game are a function of $\vec{v}$ at the previous play of the game. Therefore, we introduce the variable t to keep track of time in the game; it shall be incremented by 1 every time the game is repeated.

(4) $t \in [0, 1, 2, ..., T]$

(5) $p\alpha^t = f\alpha(\vec{v}^{\,t-1})$; $\alpha \in \{a, r, u\}$

While the payoff to each player is a function of $\vec{v}$, the payoff functions are also negotiated every so often by the user and architect. Thus, the extent to which $\vec{v}$ actually effects the payoffs to each is negotiable. The function is $f_\alpha()$ because it will be a different function for each of the architect, rogue, and user. The function $f_a()$ and $f_u()$ will produce smaller payoffs for the architect and user, respectively, with larger $\vec{v}$ since more malicious infrastructure will reduce their payoffs. $f_r()$ will produce larger payoffs for the rogue with larger $\vec{v}$. However, besides that the function is monotonically decreasing or increasing, respectively, the shape of the function (logarithmic, linear, etc.) is, in principle, negotiable.

More information about negotiated games can be found in [30]. How the payoffs are actually decided crosses into the psychology of the players and their relative power, and thus out of what pure game theory can determine. From a utility point of view the players will try to maximize their payoffs. The physical and psychological constraints of the world must be brought to bear on this negotiation modeling; otherwise it would be trivial for the players to simply set very large payoffs for everyone.

3.3.3 *Information sets*

The information available to each player will also need to be defined. In some cases, it is convenient to supply each player's subjective probability distribution over certain events for which information is incomplete [28]. However, this approach is perhaps more detailed than the present model is able to incorporate. More pertinent is each player's *information set*. The information set $\omega_\alpha^{\,t}$ for a player α is different at each point in the game t. The set $\omega_\alpha^{\,t}$ is the set of states of the game that the player knows may be the actual state of the game at time t, but between which the player cannot directly distinguish [31]. Thus, each player "knows which information set he is in, but not which vertex of the information set" [28].

Information sets help describe situations with uncertainty. In a game of 5-card poker, each player knows what cards they have, but not the cards any other player holds. However, the player knows each player has 5 cards. Certain probabilities can be calculated knowing the composition of a regular deck of cards, the player's hand, and how many people are playing. For example, if I hold 4 aces, I know that all situations in which another player holds an ace are impossible, and my information set of possible opposing hands does not include them.

Information sets in information security games are more complicated. One concrete example of this is when a user does not know whether a rogue has or has not compromised the user machine. If the user machine is infected, either the architect or the rogue could make a choice to change the user's information set. The architect can deploy accurate detection technologies and notify the user. The rogue can consume all the machine's resources, or erase the disks, which the user would notice. The user may select different strategies based on a change to their information set. Further, it does not seem that any player's information set is independent of the actions of any other player.

One element of ω_α includes a possible current state of $\vec{v}$, possible past states of $\vec{v}$, player α's past actions (in the case of imperfect recall, this will not be all past actions), as well as player α's beliefs about the possible past payoffs to the other players.

In modeling Internet security, the game's information structure is imperfect (6) and asymmetric (7), following the definitions in [31]. So there are information sets which contain more than one possible state of the game, and the information sets of different players are different. In symbols:

(6) $\forall\alpha(\exists\omega:\|\omega_\alpha\| > 1)$

(7) $\omega_\alpha{}^t = \omega_\beta{}^t \Leftrightarrow \alpha = \beta$

Exactly which elements are in $\omega_\alpha{}^t$ for each player α may be a matter of negotiation, as noted in Section 3.2. The extent the architect is permitted to monitor the user, for example, is in practice a function of the user's privacy concerns. Limiting information sets provides a formal way to discuss such concerns, as privacy partly means not being able to distinguish one user's data from another's.

4 Discussion

The game as proposed indicates some useful ways to think about the true nature of the real-world situation. The fact that the interaction between the rogue and the user is non-zero-sum is critical. This fact is due to the nature of digital resources — they are not truly rivalrous. Thus, there may be a strategy in which the rogue benefits and the user has negligible losses. In this case, the architect could not expect to impose constraints on the user to prevent the rogue's gains. The user's payoff may well be higher by not accepting such constraints. This situation helps explain the general difficulty the security community experiences with getting users to heed their warnings [3], for example.

The assertion that the game of network and Internet security as (at least) a three-person game is noteworthy. The game as described cannot be reduced to two players by putting two of the three players in a coalition. The facts of the Internet ecosystem prevent genuine coalitions in practice, and many interests of the parties do not align even in principle. Since the game has three players, a straightforward calculation of a globally-optimal strategy is not possible.

The game description also provides some practical guidance for policy and decision making. For example, if the payoff matrix is affected by the size of the rogue's infrastructure, and negotiations with the user community is stalled, then the architect's efforts would be best targeting at removing key elements of the criminal infrastructure. It also may be able to highlight certain areas that can only be solved politically as Internet governance issues, and so on.

The fact that each player has imperfect information, and that each player has different information about the game, is also a key point. Internet security is not chess, in which each player knows all the moves the other player makes — chess is a game of perfect information [30]. In chess, if one could enumerate the strategy space then one can select the globally-optimal strategy. Internet security should not be modeled as such a game, as the Internet does not function as a system with perfect information. Operational security cannot, in principle, hope to find a globally-optimal strategy.

5 Future Work

High level simulations of the posited formalisms would help to guide the plausibility of the formalisms. Establishing some hypothetical payoff matrices and attempting to calculate a solution or preferred strategy would also be an important next step. In general, all the formalizations can be made more detailed. More detail would then allow for a more rigorous analytic treatment, which would probably reveal more subtle strategic elements of the game.

The existence of any equilibria needs to be determined in order to guide other inquiries into intelligent strategies. Nash equilibria usually exist [31], for example, and a more detailed analysis could prove their existence for this game.

There is also a gap between this abstract analysis and practical measurement of the current state of affairs on the Internet that would need to be bridged before the model could be applied directly to the Internet. The present model is not sufficiently detailed to begin such measurement. Further, there is not a good framework for measuring crime on the Internet, as discussed in [5], although the authors therein propose some improvements. Eventually, such measurement efforts would need to be compatible with abstract modeling efforts so that the two can inform each other.

About the Author

Jonathan Spring is a member of the technical staff with the CERT Threat Analysis Group of the Software Engineering Institute, Carnegie Mellon University. He began working for the CERT program in 2009. He is the co-author of an information security textbook, "Introduction to Information Security: A Strategic-Based Approach," and also serves as an adjunct professor at the University of Pittsburgh's School of Information Sciences.

His research topics include monitoring cloud computing, DNS traffic analysis, and game theory. He holds a Master's degree in information security and a Bachelor's degree in philosophy from the University of Pittsburgh. Jonathan can be reached at netsa-contact@cert.org.

Acknowledgement

Thanks to Soumyo Moitra for his help in forming these ideas.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM-0000653

References

[1] : *2013 Data Breach Investigations Report (DBIR)*, 2014. URL http://www.verizonenterprise.com/DBIR/2013/.

[2] : *Black Tulip: Report of the investigation into the DigiNotar Certificate Authority breach*, 2012.

[3] Devdatta Akhawe, Adrienne Porter Felt: "Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness", *22nd USENIX Security Symposium*, 2013. URL http://www.cs.berkeley.edu/~devdatta/papers/alice-in-warningland.pdf.

[4] R. J. Anderson: *Security Engineering: A guide to building dependable distributed systems*. Wiley, 2008.

[5] R. Anderson, C. Barton, R. Böhme, R. Clayton, M.J.G. van Eeten, M. Levi, T. Moore, S. Savage: "Measuring the cost of cybercrime", *11th Workshop on the Economics of Information Security*, 2012. URL http://weis2012.econinfosec.org/papers/Anderson_WEIS2012.pdf.

[6] Steven J Brams: *Negotiation Games: Applying game theory to bargaining and arbitration*. Routledge, 2003.

[7] Huseyin Cavusoglu, Birendra Mishra, Srinivasan Raghunathan: "A model for evaluating IT security investments", *Communications of the ACM*, pp. 87—92, 2004.

[8] Adam Cummings, Todd Lewellen, David McIntire, Andrew Moore, Randall Trzeciak: *Insider Threat Study: Illicit Cyber Activity Involving Fraud in the U.S. Financial Services Sector*, 2012. URL http://www.sei.cmu.edu/library/abstracts/reports/12sr004.cfm.

[9] David Drummond: *A new approach to China*. Google Official Blog, 2010.

[10] L. Dolanskỳ: "Present state of the Lanchester theory of combat", *Operations Research*, pp. 344—358, 1964.

[11] Ellen Messmer: "RSA's SecurID security breach: What should you do?", *Network World*, 2011. URL http://www.networkworld.com/news/2011/031811-rsa-securid-breach.html.

[12] Ellen Messmer: "RSA's SecurID security breach: What should you do?", *Network World*, 2011. URL http://www.networkworld.com/news/2011/031811-rsa-securid-breach.html.

[13] Drew Fudenberg, Jean Tirole: *Game theory*. 1991. MIT Press, 1991.

[14] Herbert Gintis: *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton University Press, 2000.

[15] Kuno JM Huisman: *Technology Investment: a game theoretic real options approach*. Kluwer Academic Pub, 2001.

[16] John Gilmore: DES (*Data Encryption Standard*) Review at Stanford University, 2005. URL http://www.toad.com/des-stanford-meeting.html.

[17] C. Kanich, N. Weaver, D. McCoy, T. Halvorson, C. Kreibich, K. Levchenko, V. Paxson, G.M. Voelker, S. Savage: "Show Me the Money: Characterizing Spam-advertised Revenue", *20th USENIX Security Symposium*, 2011. URL https://www.usenix.org/legacy/event/sec11/tech/full_papers/Kanich.pdf.

[18] Ioanna Kantzavelou, Sokratis Katsikas: "A game-based intrusion detection mechanism to confront internal attackers", *Computers & Security*, pp. 859—874, 2010.

[19] MK Lauren: *Describing Rates of Interaction between Multiple Autonomous Entities: An Example Using Combat Modelling*, 2001.

[20] S.D. Moitra: *Managing Risk from Cybercrime: Internet Policy and Security Management for Organizations*. Max-Planck-Institut f. ausländisches und internationales Strafrecht, 2008.

[21] Tyler Moore, Richard Clayton: "Evil searching: Compromise and recompromise of internet hosts for phishing", *Financial Cryptography and Data Security*, pp. 256—272, 2009.

[22] Roger B Myerson: *Game theory: analysis of conflict*. Harvard University Press, 1997.

[23] John F Nash Jr: "Non-cooperative games", *The Annals of Mathematics*, pp. 286—295, 1951.

[24] John F Nash Jr: "The bargaining problem", *Econometrica: Journal of the Econometric Society*, pp. 155—162, 1950.

[25] G. Owen: *Game theory*. Emerald Group Publishing, 1995.

[26] Anatol Rapoport: *N-person game theory: Concepts and applications*. Courier Dover Publications, 1970.

[27] Anatol Rapoport: *Two-person game theory: The essential ideas*. Courier Dover Publications, 1966.

[28] E. Rasmusen: *Games and Information: An Introduction to Game Theory*. Blackwell, 2007.

[29] R. Rasmussen, G. Aaron: *Global phishing survey: trends and domain name use in 2Q2012*, 2012.

[30] Sankardas Roy, Charles Ellis, Sajjan Shiva, Dipankar Dasgupta, Vivek Shandilya, Qishi Wu: "A survey of game theory as applied to network security", *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pp. 1—10, 2010.

[31] J.M. Spring: "Modeling Malicious Domain Name Take-down Dynamics: Why eCrime Pays", *IEEE eCrime Researchers Summit*, 2013. URL http://resources.sei.cmu.edu/library/asset-view.cfm?assetID=88265

[32] T Spyridopoulos, G Karanikas, T Tryfonas, G Oikonomoug: "A Game Theoretic Defence Framework Against DoS/DDoS Cyber Attacks", *Computers & Security*, pp. 39—50, 2013.

[33] John Von Neumann, Oskar Morgenstern: *The theory of games and economic behavior*. Princeton university press, 1944.

[34] E Weinan, Bjorn Engquist, Xiantao Li, Weiqing Ren, Eric Vanden-Eijnden: "Heterogeneous multiscale methods: a review", *Communications in computational physics*, pp. 367—450, 2007.

[35] William Casey, Jose A. Morales, Thomson Nguyen, Jonathan Spring, Rhiannon Weaver, Evan Wright, Leigh Metcalf, Bud Mishra: "Cyber Security via Signaling Games: Toward a Science of Cyber Security", *ICDCIT*, pp. 34-42, 2014. URL http://dx.doi.org/10.1007/978-3-319-04483-5_4.

[36] Quanyan Zhu, Linda Bushnell, Tamer Basar: "Game-theoretic analysis of node capture and cloning attack with multiple attackers in wireless sensor networks", *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 3404—3411, 2012.

Cyber Profiling: Using Instant Messaging Author Writeprints for Cybercrime Investigations

By Angela Orebaugh, Jason Kinser, and Jeremy Allnutt

The explosive growth in the use of instant messaging (IM) communication in both personal and professional environments has resulted in an increased risk to proprietary, sensitive, and personal information and safety due to the influx of IM-assisted cybercrimes, such as phishing, social engineering, threatening, cyber bullying, hate speech and crimes, child exploitation, sexual harassment, and illegal sales and distribution of software. IM-assisted cybercrimes are continuing to make the news with child exploitation, cyber bullying, and scamming leading last month's headlines. Instant messaging's anonymity and use of virtual identities hinders social accountability and presents a critical challenge for cybercrime investigation. Cyber forensic techniques are needed to assist cybercrime decision support tools in collecting and analyzing digital evidence, discovering characteristics about the cyber criminal, and assisting in identifying cyber criminal suspects.

Introduction

The anonymous nature of the Internet allows online criminals to use virtual identities to hide their true identity to facilitate cybercrimes. Although central IM servers authenticate users upon login, there is no means of authenticating or validating peers (buddies). Current IM products are not addressing the anonymity and ease of impersonation over instant messaging. Author writeprints can provide cybercrime investigators a unique tool for analyzing IM-assisted cybercrimes. Writeprints are based on behavioral biometrics, which are persistent personal traits and patterns of behavior that may be collected and analyzed to aid a cybercrime investigation. (Li et al., 2006) Instant messaging behavioral biometrics include online writing habits, known as stylometric features, which may be used to create an author writeprint to assist in identifying an author, or characteristics of an author, of a set of instant messages. The writeprint is a digital fingerprint that represents an author's distinguishing stylometric features that occur in his/her computer-mediated communications. Writeprints may be used as input to a criminal cyberprofile and as an element of a multimodal system for cybercrime investigations. Writeprints can be used in conjunction with other evidence, criminal investigation techniques, and biometrics techniques to reduce the potential suspect space to a certain subset of suspects; identify the most plausible author of an IM conversation from a group of suspects; link related crimes; develop an interview and interrogation strategy; and gather convincing digital evidence to justify search and seizure and provide probable cause.

Instant Messaging and Cybercrime

Instant messaging's anonymity hinders social accountability and leads to IM-assisted cybercrime facilitated by the following:

- Users can create any virtual identity,
- Users can log in from anywhere,
- Files can be transmitted, and
- Communication is often transmitted unencrypted.

In IM communications, criminals use virtual identities to hide their true identity. They can use multiple screen names or impersonate other users with the intention of harassing or deceiving unsuspecting victims. Criminals may also supply false information on their virtual identities, for example a male user may configure his virtual identity to appear as female. Since most IM systems use the public Internet, the risk is high that usernames and passwords may be intercepted, or an attacker may hijack a connection or launch a man-in-the-middle (MITM) attack. With hijacking and MITM attacks, the victim user thinks he/she is communicating with a buddy but is really communicating with the attacker masquerading as the victim's buddy. Instant messaging's anonymity allows cyber criminals such as pedophiles, scam artists, and stalkers to make contact with their victims and get to know those they target for their crimes (Cross, 2008). IM-assisted cybercrimes, such as phishing, social engineering, threatening, cyber bullying, hate speech and crimes, child exploitation, sexual harassment, and illegal sales and distribution of software are continuing to increase (Moores and Dhillon, 2000). Additionally, criminals such as terrorist groups, gangs, and cyber intruders use IM to communicate (Abbasi and Chen, 2005). Criminals also use IM to transmit worms, viruses, Trojan horses, and other malware over the Internet.

With increasing IM cybercrime, there is a growing need for techniques to assist in identifying online criminal suspects as part of the criminal investigation. Cyber forensics is the application of investigation and analysis techniques to gather evidence suitable for presentation in a court of law with the goal of discovering the crime that took place and who was responsible (Bassett et al., 2006). With IM communications, it is necessary to have cyber forensics techniques to assist in determining the IM user's real identity and collect digital evidence for investigators and law enforcement.

Behavioral Biometrics Writeprints for Authorship Analysis

Determining an IM user's real identity relies on the fact that humans are creatures of habit and have certain persistent personal traits and patterns of behavior, known as behavioral biometrics (Revett, 2008). Online writing habits, known as stylometric features, include composition syntax and layout, vocabulary patterns, unique language usage, and other stylistic traits. Thus, certain stylometric features may be used to create an author writeprint to help identify an author of a particular piece of work (De Vel et al., 2001). A writeprint represents an author's distinguishing stylometric features that occur in his/her instant messaging communications. These stylometric features may include average word length, use of punctuation and special characters, use of abbreviations, and other stylistic traits. Writeprints can provide cybercrime investigators a unique behavioral biometric tool for analyzing IM-assisted cybercrimes. Writeprints can be used as input to a criminal cyberprofile and as an element of a multimodal system to perform cyber forensics and cybercrime investigations.

Instant messaging communications contain several stylometric features for authorship analysis research. Certain IM specific features such as message structure, unusual language usage, and special stylistic markers are useful in forming a suitable writeprint feature set for authorship analysis (Zheng et al., 2006). The style of IM messages is very different than that of any other text used in traditional literature or other forms of computer-mediated communication. The real time, casual nature of IM messages produces text that is conversational in style and reflects the author's true writing style and vocabulary (Kucukyilmaz et al., 2008). Significant characteristics of IM are the use of special linguistic elements such as abbreviations, and computer and Internet terms, known as netlingo. The textual nature of IM also creates a need to exhibit emotions. Emotion icons, called emoticons, are sequences of punctuation marks commonly used to represent feelings within computer-mediated text (Kucukyilmaz et al., 2008). An author's IM writeprint may be derived from network packet captures or application data logged during an instant messaging conversation. Although some types of digital evidence, such as source IP addresses, file timestamps, and metadata may be easily manipulated, author writeprints based on behavioral biometrics are unique to an individual and difficult to imitate.

Creating IM Writeprints

A stylometric feature set is composed of a predefined set of measurable writing style attributes. Given t predefined features, each set of IM messages for a given author can be represented as a t-dimensional vector, called a writeprint. Figure 1 presents a stylometric feature set for a 356-dimensional vector writeprint with lexical, syntactic, and structural features. (Orebaugh et al., 2014) The number of features in each category is shown in parenthesis.

Lexical features mainly consist of count totals and are further broken down into emoticons, abbreviations, word-based, and character-based features. Syntactic features include punctuation and function words in order to capture an author's habits of organizing sentences. Function words include conjunctions, prepositions, and other words that carry little meaning when used alone, such as "the" or "of". They

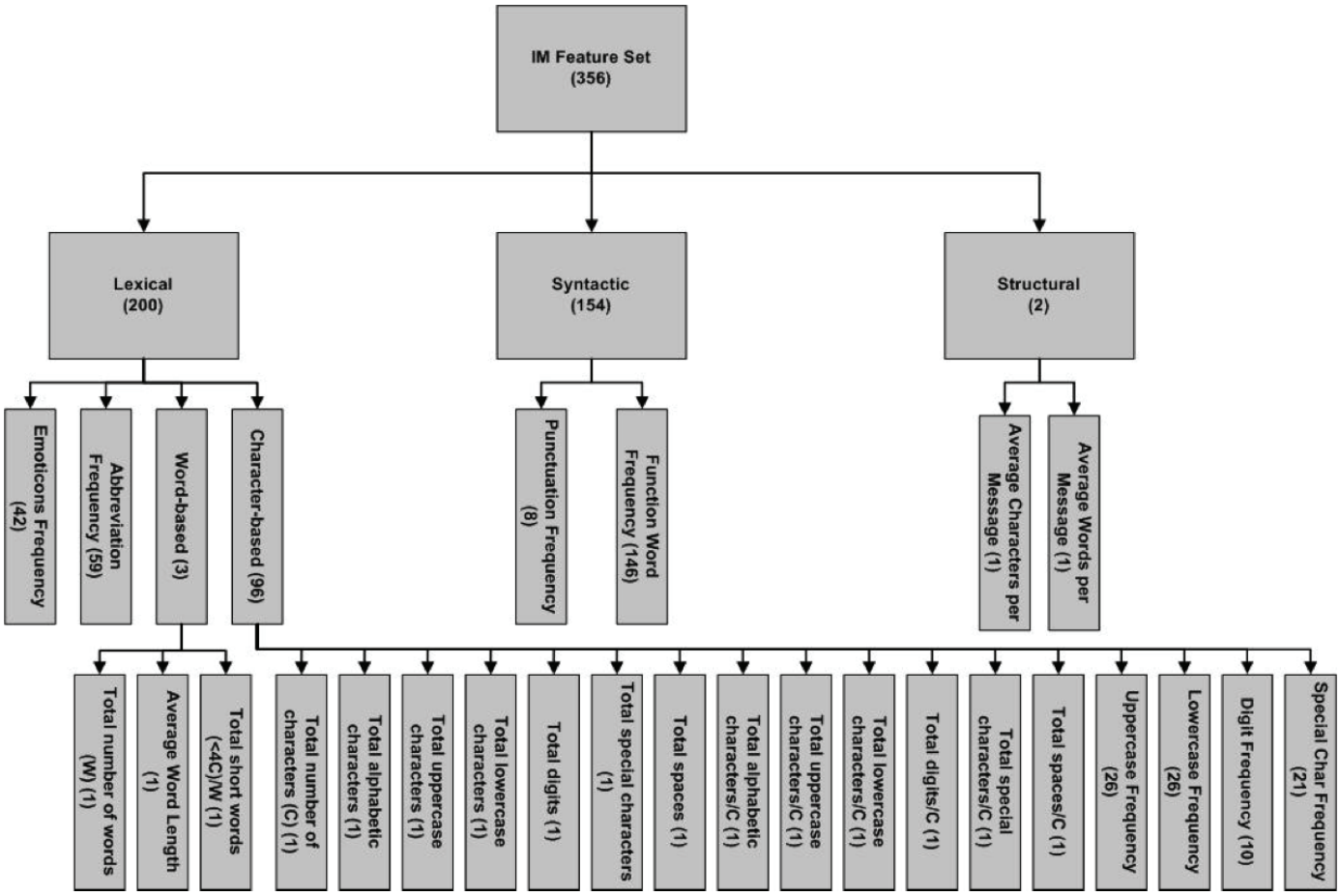


Figure 1. IM Writeprint Feature Set

provide relationships to content words in the sentence, such as "ball" or "bounce". Analyzing function words as opposed to content words allows topic-independent results that reflect an author's preferred ways to express himself or herself and form sentences. Structural features capture the way an author organizes the layout of text. With IM communications there are no standard headers, greetings, farewells, or signatures, leaving simply the average characters and words per message in terms of structural layout. A list of function words, abbreviations, and emoticons are included in Appendix A.

Writeprints are created by generating totals for each stylometric feature, resulting in the output of a writeprint (W_x) for a set of messages $\{M_1,\ldots,M_p\}$ for an author (A_n) or author category (C_m). A writeprint may be viewed in a comma-separated value (CSV) format where each value represents a total for a specific feature. An example writeprint for an author A_n using a selected feature set $\{F_1,\ldots,F_q\}$, where q =100, for a set of messages $\{M_1,\ldots,M_p\}$ looks like the following:

```
105,1,0,0,4,0,1250,0,4,0,18,
8,1,2,0,0,0,0,1,9,0,14,31,6.
78,3.71,23,0,67,4,25,5,0,117
,5,0,1,4,0,0,23,0,0,0,8,0,23
,1,3,0,27,50,0,0,1550,0,7,0,
0,0,1,0,1250,33,0,13,1,0,0,0
,2,85,0,0,0,4,0,0,0,0,0,96,1
,0,0,0,13,0,3,0,10,0,2,0,0,0
,1,2,16,0,0.806
```

After writeprints are generated they may then be normalized, standardized, and input into various statistical models for analysis. Figure 2 shows the output of the Principal Component Analysis (PCA) model for writeprints for seven authors. (Orebaugh et al., 2014) The figure shows the first 3 principal components for multiple author conversations, mapped in three-dimensional space. In this example, each author has a relatively well-defined cluster representing his or her writeprint. Different authors separate from each other, while multiple conversations of an author cluster together.

This type of example may be used in an investigation to show that sample evidentiary writeprints do or do not overlap with certain suspect writperints, thus helping investigators narrow the suspect space, develop an interrogation strategy, link related crimes, or justify probable cause.

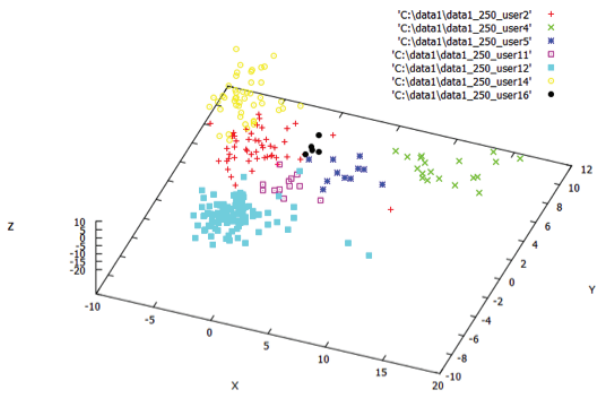


Figure 2. IM Writeprint PCA Output

Cybercrime Investigations and IM

Many disciplines including psychology, philosophy, sociology, criminology, law, knowledge management, and computer science have studied the criminal investigation process. Although cybercrime is a relatively new form of crime that has rapidly evolved over the last few decades, cybercrime investigations and traditional criminal investigations share the same goal – to gather information. Figure 3 illustrates the traditional criminal investigation process as presented in *Scene of the Cybercrime* (Cross, 2008).

The investigator first determines if an act has violated the law and warrants an investigation. Next, evidence is collected and analyzed, including tangible evidence such as hard drives and electronic devices, and the digital evidence they contain. Cybercrime investigations for IM rely on instant messaging exchanges, or conversations, as digital evidence. The sources for IM digital evidence include both data and meta-data. The data includes the IM text and the meta-data includes other related evidence such as the IM client version, timestamps, the length of time the user has been logged on, etc. The next step involves seeking expert advice if necessary. Often times in cybercrime cases the investigator needs to seek expert advice on the technical aspects of the crime. Experts may be on staff, or may be located from professional organizations, consultants, or the academic community. For IM related cybercrimes expert witnesses may include linguists, communication experts, or social psychologists. The next step of interviewing witnesses and interrogating suspects is an ongoing process throughout the investigation as new witnesses and suspects are discovered. Throughout this stage suspects are eliminated and the most plausible suspect is identified. Next, the investigator begins preparing the case file to include the initial incident report, evidence, other reports such as lab reports, written statements, and other relevant information. Once the case

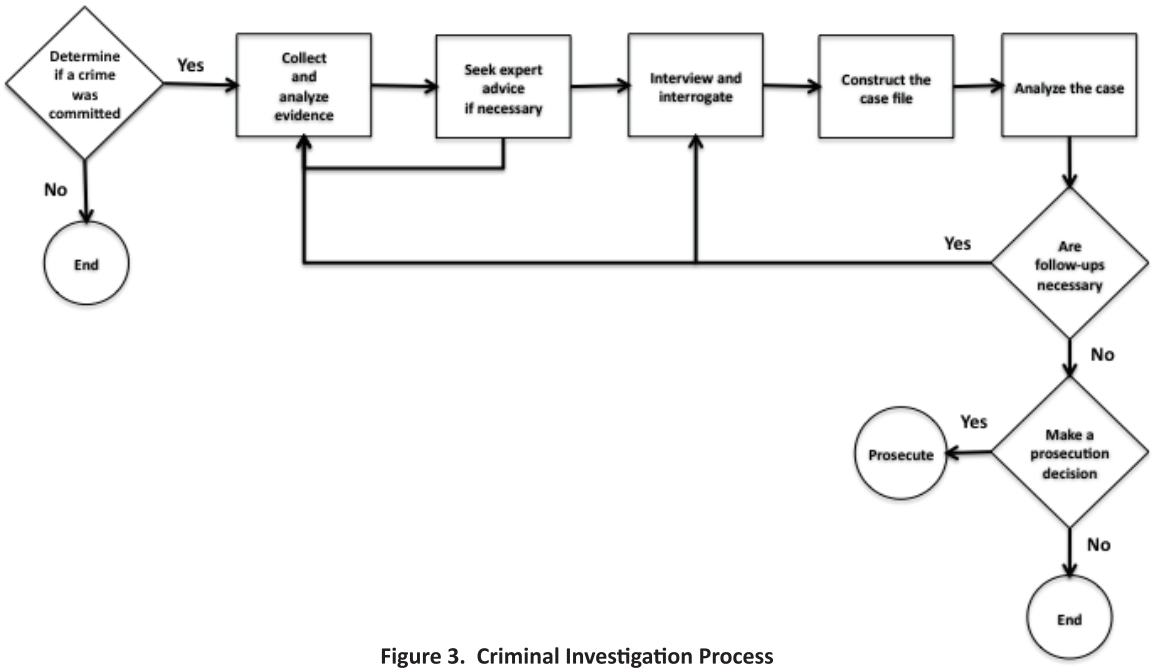


Figure 3. Criminal Investigation Process

file is constructed it is analyzed to determine weaknesses and to identify additional information needed for prosecution. This analysis leads to any follow-up investigations that need to occur including collecting additional evidence and interviewing witnesses again. Once the case is considered complete the prosecutor will decide whether to bring the case to trial and how to proceed. There is no standard accuracy measure or probability threshold for authorship attribution evidence; the investigator only needs probable cause to initiate a warrant or arrest. In addition, evidence admissibility varies by jurisdiction. In cases where digital evidence is not admissible, expert witnesses are often called upon to provide their expertise and interpretation. In the court of law, the jury only needs reasonable doubt to determine a defendant's guilt or innocence. Some relevant criminal cases were investigated and prosecuted based on text message abbreviations, sentence length, and punctuation (Leafe, 2009).

Criminal Profiling and IM

Criminal profiling is an investigative method that has been used in traditional criminal investigations that can also be applied to cybercrime investigations, known as cyberprofiling. Cross defines traditional criminal profiling is the "art and science of developing a description of a criminal's characteristics (physical, intellectual, and emotional) based on information collected at the scene of the crime" (Cross, 2008). Criminal profiling often uses patterns and correlations among criminal activity and different crimes to construct a profile. Criminal profiling is used to assist with the investigative process, reduce the potential suspect space to a certain subset of suspects, link related crimes, and develop an interview and interrogation strategy (Casey, 1999). It is important to note that a criminal profile will only provide generalities about the type of person who committed a crime, it will not identify a specific individual. Criminal profiling is one method among many for assisting with criminal investigations and building a case file. The profile cannot exist as evidence, rather it provides information to allow investigators to focus on the right suspects and begin to gather additional evidence (Cross, 2008). A criminal profile can be used in court in conjunction with expert witness testimony. "An expert witness can reference a criminal profile as the basis of an opinion that there is a high probability of a link between a particular suspect and a particular crime" (Cross, 2008). An IM author writeprint may be used as input to a criminal profile.

The FBI is credited with formalizing the criminal profiling process. The FBI's Behavioral Science Unit (BSU) "focuses on developing new and innovative investigative approaches and techniques to solve crimes by studying offenders and their behaviors and motivations" (FBI, 2014). The BSU has been assisting local, state, and federal agencies in narrowing investigations by providing criminal profiles since the 1970s (Doublas et al., 2014). The FBI BSU has created the six-step criminal profile generating process shown in Table 1.

Table 1. FBI BSU Criminal Profile Process

FBI BSU Criminal Profile Process	
1. Profiling Inputs	The first step collects profiling inputs including comprehensive information about the crime and all evidence collected, both tangible, physical evidence and digital evidence.
2. Decision Process Models	This step analyzes the information and evidence to determine patterns and possible linkages to other crimes.
3. Crime Assessment	The crime scene is reconstructed and analyzed to determine the sequence of events and other information about the crime.
4. Criminal Profile	The first three steps are combined to create a criminal profile, often incorporating the motives, physical qualities, and personality of the perpetrator. The criminal profile is also used to create an interrogation strategy for the suspects.
5. The Investigation	Investigators and others use the profile to learn more information and identify suspects. Suspects matching the profile are evaluated. The profile may be reassessed if no leads or suspects are identified.
6. The Apprehension	The last stage occurs when investigators believe they have identified the most plausible suspect likely to be the perpetrator. A warrant is obtained for the arrest of the individual, usually followed by a trial (Doublas et al., 2014).

The FBI criminal profile generating process may be easily applied in a cybercrime investigation to perform cyberprofiling. Various types of digital and non-digital evidence may be combined as profile inputs, including, email, IM conversations, network packet captures, account activity information, and physical evidence. A

cybercriminal's profile may include a number of traits such as time and location of computer access, types of computer attacks launched by the attacker, programs and attack tools used, writeprints, and targets of the cybercrime whether they be human or electronic (networks, satellites, phones, computer systems, etc.).

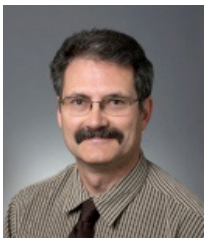
In the context of IM-assisted cybercrime, cyberprofiling uses IM data such as the conversation logs, IM client version, timestamps, the length of time the user has been logged on, etc. IM writeprints may be used in conjunction with other evidence and investigative techniques to build or validate a criminal profile; reduce the potential suspect space to a certain subset of suspects; link related crimes; develop an interview and interrogation strategy; and gather convincing digital evidence to justify search and seizure and provide probable cause.

Conclusion

As cybercrimes continue to increase, new cyber forensics techniques are needed to combat the constant challenge of Internet anonymity. The IM writeprint technique may be used to assist cybercrime decision support tools in collecting and analyzing digital evidence, discovering characteristics about the cyber criminal, and assisting in identifying cyber criminal suspects. Future areas of research include implementing the IM writeprint taxonomy on past and/or ongoing investigation data for further analysis and modification. Additionally, this research would benefit from a feasibility analysis of various sociolinguistic writeprint categories (such as gender and age). Lastly, the IM writeprint taxonomy may be modified and applied to other communication mediums such as text, Twitter, and Facebook.

About the Authors

Dr. Angela Orebaugh is Fellow and Chief Scientist at Booz Allen Hamilton. She received her Ph.D. from George Mason University with a concentration in Information Security. Her current research interests include behavioral biometrics and the Internet of Things.



Dr. Jason Kinser is an Associate Professor in the School of Physics, Astronomy, and Computational Sciences at George Mason University. His current research interests include classification of regions in lung scans to detect idiopathic pulmonary fibrosis.

Dr. Jeremy Allnutt is a Professor in the Electrical and Computer Engineering Department at George Mason University with a focus in communications and signal processing, computer networking, and telecommunications.

References

1. Cross, Michael. *Scene of the Cybercrime*. Syngress Publishing, (2008): 679-690
2. Moores, Trevor, and Gurpreet Dhillon. "Software piracy: a view from Hong Kong." *Communications of the ACM* 43.12 (2000): 88-93.
3. Abbasi, Ahmed, and Hsinchun Chen. "Applying authorship analysis to extremist-group web forum messages." *Intelligent Systems*, IEEE 20.5 (2005): 67-75.
4. Bassett, Richard, Linda Bass, and Paul O'Brien. "Computer forensics: An essential ingredient for cyber security." *Journal of Information Science and Technology* 3.1 (2006): 22-32.
5. Revett, Kenneth. *Behavioral biometrics: a remote access approach*. Wiley Publishing, (2008): 1-2.
6. De Vel, Olivier, Alison Anderson, Malcolm Corney, and George Mohay. "Mining e-mail content for author identification forensics." *ACM Sigmod Record* 30.4 (2001): 55-64.
7. Zheng, Rong, Jiexun Li, Hsinchun Chen, and Zan Huang. "A framework for authorship identification of online messages: Writing-style features and classification techniques." *Journal of the American Society for Information Science and Technology* 57.3 (2006): 378-393.
8. Kucukyilmaz, Tayfun, B. Cambazoglu, Cevdet Aykanat, and Fazli Can. "Chat mining: Predicting user and message attributes in computer-mediated communication." *Information Processing & Management* 44.4 (2008): 1448-1466.
9. Leafe, David. "Dear Garry. I've decided to end it all: The full stop that trapped a killer." *Daily Mail* (2009).
10. Casey, E. "Cyberpatterns: criminal behavior on the Internet." *Criminal profiling: An introduction to behavioral evidence analysis* (1999): 361-378.
11. Federal Bureau of Investigation, Behavioral Science Unit website. http://www.fbi.gov/hq/td/academy/bsu/bsu.htm. (accessed March 4, 2014)
12. Doublas, John E., Robert K. Ressler, Ann W. Burgess, and Carol R. Hartman. "Criminal profiling from crime scene analysis." *Behavioral Sciences & the Law* 4.4 (1986): 401-421.
13. Li, Jiexun, Rong Sheng, and Hsinchun Chen. "From Fingerprint to Writeprint." *Communications of the ACM* 49.4 (2006): 76-82
14. Orebaugh, Angela, Jason Kinser, and Jeremy Allnutt. "Visualizing Instant Messaging Author Writeprints for Forensic Analysis," In Proceedings of Conference on Digital Forensics, Security and Law, Richmond VA (2014): 191-213

Appendix A

Function Words					
about	both	inside	of	something	we
above	but	into	off	such	what
after	by	is	on	than	whatever
all	can	it	once	that	when
although	could	its	onto	the	where
am	do	latter	opposite	their	whether
among	down	less	or	them	which
an	each	like	our	these	while
and	either	little	outside	they	who
another	enough	lots	over	this	whoever
any	every	many	own	those	whom
anybody	everybody	me	past	though	whose
anyone	everyone	more	per	through	will
anything	everything	most	plenty	till	with
are	few	much	plus	to	within
around	following	must	regarding	towards	without
as	for	my	same	under	worth
at	from	near	several	unless	would
be	have	need	she	unlike	yes
because	he	neither	should	until	you
before	her	no	since	up	your
behind	him	nobody	so	upon	
below	if	none	some	us	
beside	in	nor	somebody	used	
between	including	nothing	someone	via	
Abbreviations					
143	CYA	ILY	OMG	THX	WYWH
…	DBEYR	IMHO	OTP	TLC	XOXO
2moro	DILLIGAS	IRL	PITA	TMI	YT
2nite	ETC	ISO	PLS	TTYL	YW
ASAP	FUBAR	JK	PLZ	TTYS	
B4N	FWIW	L8R	POV	TYVM	
BCNU	FYI	LMAO	ROTFL	U2	
BFF	GR8	LMFAO	RU	VBG	
BRB	IC	LOL	SOL	WEG	
BTW	IDC	NP	STBY	WTF	
CU	IDK	OIC	SWAK	WTG	
Emoticons					
:-)	:)	:-(	:(	;-)	;)
:-P	:P	;-P	;P	:-D	:D
:'-(	:'(	:*	:-*	0:-)	0;-)
:-!	:*(	>:)	>:-)	:-*	:-/
:-\	:-[	:-]	:-{	:-}	:-S
:-x	:-#	:-|	=)	>:-(	>:(
<3	</3	0:)	:*	:/	:\

BECO: Behavioral Economics of Cyberspace Operations

By Victoria Fineberg

This paper proposes a risk-management framework Behavioral Economics of Cyberspace Operations (BECO) for hardening Cyberspace Operations (CO) with the Behavioral Economics (BE) models of cognitive biases in judgment and decision-making. In applying BE to CO, BECO augments a common assumption of a rational cyber warrior with more realistic expressions of human behavior in cyberspace. While the current development of the cyber workforce emphasizes education and training, BECO addresses typical conditions under which rational decision-making fails and knowledge is neglected. The BECO framework encompasses a full set of cyber actors, including attackers, defenders, and users on the friendly and adversary sides, across the full CO spectrum in space and time, and offers a structured approach to the cognitive bias mitigation.

Bringing BE into CO

This paper proposes enhancements of Cyberspace Operations (CO) by adapting Behavioral Economics (BE) models in a novel framework Behavioral Economics of Cyberspace Operations (BECO). The essence of BECO is the identification of cognitive biases of CO actors, mitigation of biases on the friendly side, and exploitation of biases on the adversary side. BECO is a CO-focused extension of the Behavioral Economics of Cybersecurity (BEC) framework (Fineberg, 2014) that augments the National Institute of Standards and Technology's Risk Management Framework (RMF) of information security (NIST SP 800-39, 2011) by introducing a new class of vulnerabilities corresponding to persistent human biases. BECO takes it further by applying the BEC risk management approach to cyber operations and CO-specific cyberactors. Figure 1 depicts the progression from BE to BEC and BECO and the concepts that link them.

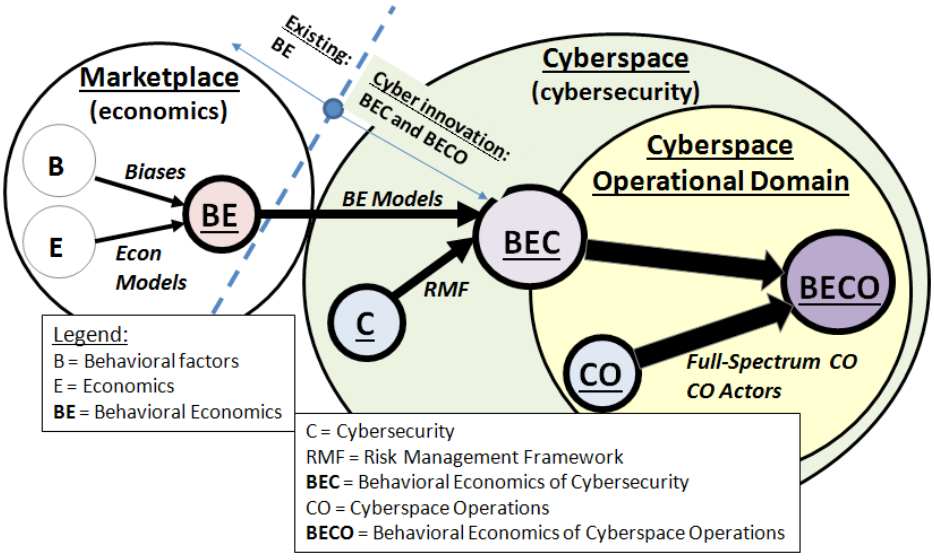


Figure 1. Progression from BE to BEC and BECO.

While the current cognitive analysis of warfighting is rooted in psychology (Grossman and Christensen, 2007), the awareness of the BE discoveries is rising in the military community (Mackay & Tatham, 2011; Holton, 2011). However, in the existing work, the BE relevance is limited to providing general analogies between the BE findings and military scenarios, without offering a practical approach for using BE in the operations. In contrast, BECO provides an overarching framework of behavioral models encompassing the full spectrum of

operational scenarios and cyberactors. The goals of this work are to raise the awareness of persistent human biases of CO actors that cannot be eliminated by traditional training, provide a framework for identifying and mitigating critical biases, and influence policies guiding cyberspace security and operations.

Cyberspace Operations and BECO

The CO concept is evolving, and this paper uses the current tenets of the United States Cyber Command (USCYBERCOM) as the basis for analyzing the CO characteristics addressed in BECO. CO are conducted in cyberspace, which Department of Defense (DoD) has designated as a warfighting domain (Stavridis & Parker, 2012, p. 62) and a part of the Information Environment (IE) that exists in three dimensions: Physical, Informational, and Cognitive. CO is a component of the Information Operations (IO) conducted in IE, as shown in Figure 2.

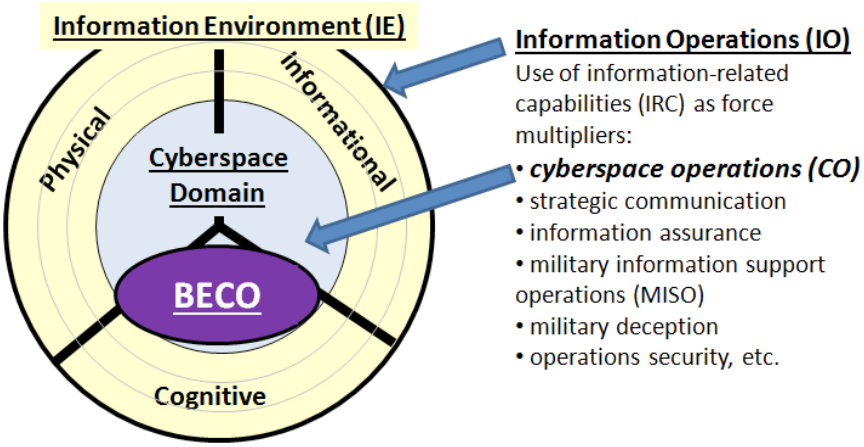


Figure 2. Key concepts related to Cyberspace Operations.

The joint doctrine defines the *Information Environment (IE)* as "the aggregate of individuals, organizations, and systems that collect, process, disseminate, or act on information" (JP 3-13, 2012, p. vii); the *Information Operations (IO)* as "the integrated employment, during military operations, of [Information Related Capabilities] IRCs in concert with other lines of operation to influence, disrupt, corrupt, or usurp the decision making of adversaries and potential adversaries while protecting our own" (p. vii); *Cyberspace* as "a global domain within the information environment consisting of the interdependent network of information technology infrastructures and resident data, including the Internet, telecommunications networks, computer systems, and embedded processors and controllers" (p. II-9); and the *Cyberspace Operations (CO)* as "the employment of cyberspace capabilities where the primary purpose is to achieve objectives in or through cyberspace" (p. II-9). The IE migration towards the Joint Information Environment (JIE) will facilitate the cyberspace defense, and BECO will enhance JIE's cognitive dimension.

The USCYBERCOM's mission is to conduct the full-spectrum CO in the three focus areas including the defense of the DoD Information Networks (DoDIN), support of combatant commanders, and response to cyber attacks (U.S. Cyber Command, 2013). Correspondingly, USCYBERCOM operates across three Lines Of Operation (LOO) including DoD Network Operations (DNO), Defensive Cyber Operations (DCO), and Offensive Cyber Operations (OCO) (Pellerin, 2013a). DNO provides a static defense of the DoDIN perimeter. DCO includes maneuvers within the perimeter to stop attacks that have passed the static DNO defenses, actions outside the perimeter to stop impending attacks, and employment of Red Teams. OCO is "the ability to deliver a variety of effects outside our own network to satisfy national security requirements" (Pellerin, 2013a). Figure 3 below provides a graphical representation[1] of these COs.

BECO uses the full-spectrum nature of USCYBERCOM to define a comprehensive set of cognitive CO scenarios, as discussed below.

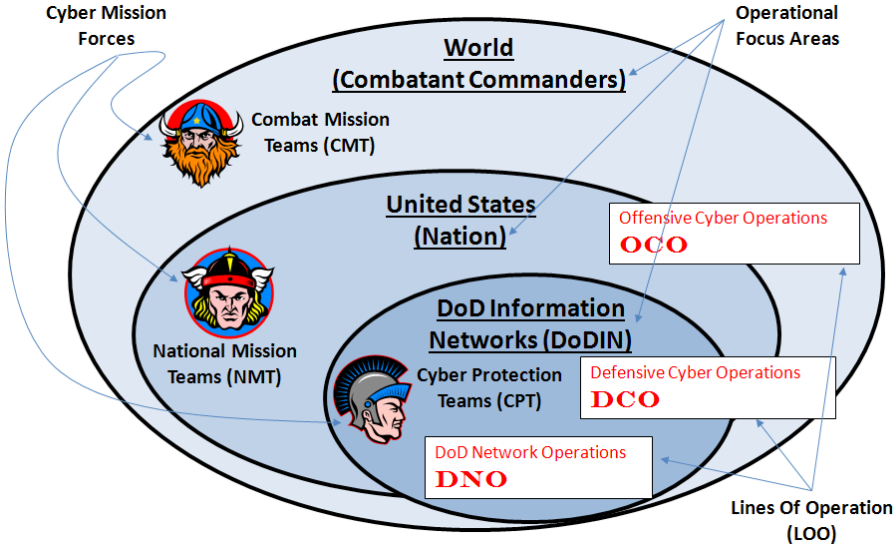


Figure 3. Graphical representation of the CYBERCOM COs.

1 This figure is developed for this paper as a graphical representation of the CYBERCOM COs using publically available information. The figure is not developed or endorsed by the CYBERCOM and is used for illustration only.

Behavioral Economics

This section provides some BE background with the emphasis on the BECO relevance.

BE Background

Behavioral Economics (BE) is a recent science that emerged at the confluence of psychology and economics to correct Standard Economics (SE) models for cognitive biases demonstrated in psychological experiments. SE relies on the *rational-agent* model of the preference-maximizing human behavior. In contrast, BE is based on the statistically significant evidence of systematic deviations of the economic actors' behavior from the rationality assumed in SE. Economists use the terms 'rationality' and 'biases' in a specific context. Kahneman, a 2002 winner of the Nobel Memorial Prize in Economic Sciences, explains that *rationality* is logical coherence, which could be reasonable or not (2011). The rational-agent model assumes that people use information optimally and that the cost of thinking is constant. However, empirical evidence shows that even high-stake strategic decisions are biased (Kahneman, 2013). A *bias* is a systematic error, an average system error that is different from zero (Kahneman, 2006). BE studies biases that represent psychological mechanisms skewing people's decisions in specific directions, beyond the considerations of rationality and prudence.

Psychology: Fast and Slow Thinking

The differences between biased and rational decision making can be traced to the distinction between two types of thinking that Kahneman (2011) calls System 1 (S1) and System 2 (S2), respectively. S1 refers to the fast, automatic, intuitive thinking; and S2 refers to the slow, deliberate, effortful thinking. The S1 thinking includes *automatic* activities of memory and perception; and *intuitive* thoughts of two types, the expert and the heuristic. The *expert* thought is fast due to prolonged practice, and the *heuristic* thought is exemplified by one's ability to complete the phrase 'bread and …' and answer 2 + 2 = ? In contrast with S1, S2 performs effortful mental activities that require concentration. Examples of S2 activities include parking a car in a narrow space, filling out tax forms, and complex computations. Figure 4 summarizes the key features of S1 and S2 with the emphasis on the S1-based heuristics that are the main cause of cognitive biases in judgment and decision making.

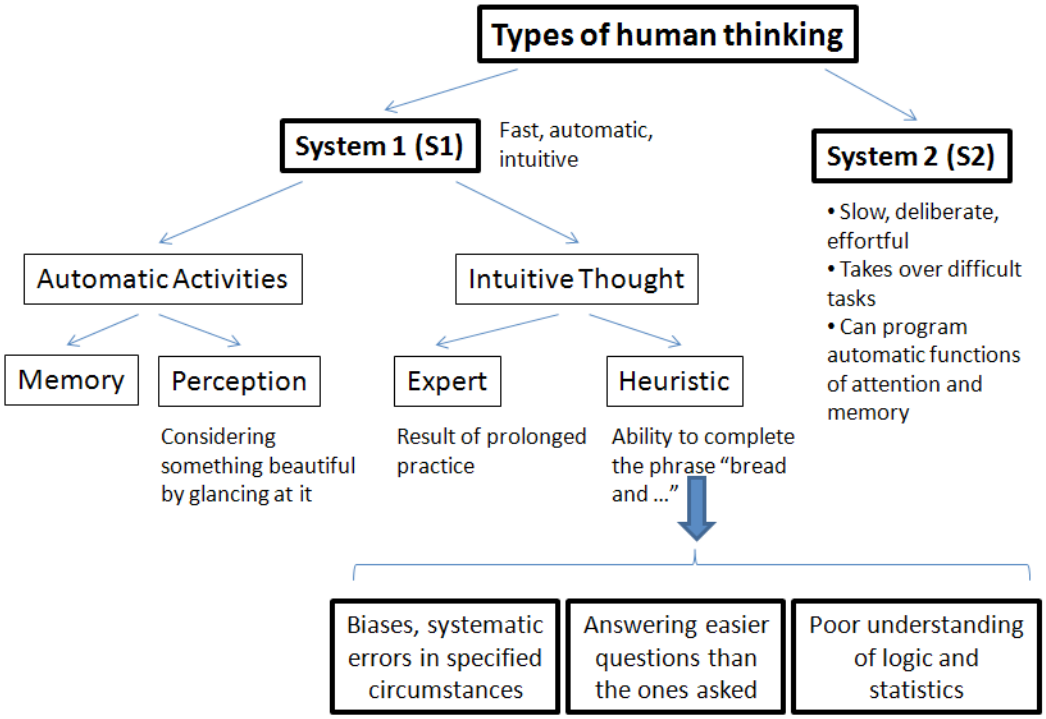


Figure 4. Types of human thinking.

Interactions between S1 and S2 are complex and generally favor decisions made by S1, even though S2 has some limited capacity to program normally-automatic functions of attention and memory. S1 produces biases, which are systematic errors it makes in specific circumstances, such as answering easier questions than those asked and misunderstanding logic and statistics.

S2 is used to focus on a task, but the intense focus blinds people to other stimuli and cannot be sustained for prolonged periods of time. Most thinking originates in S1, but S2 takes over when decisions are difficult and has the last word. While it may be desirable to switch from S1 to S2 in order to avoid making biased choices, Kahneman notes that "because System 1 operates automatically and cannot be turned off at will, errors of intuitive thought are often difficult to prevent. Biases cannot always be avoided, because System 2 may have no clue to the error. Even when cues to likely errors are available, errors can be prevented only by the enhanced monitoring and effortful activity of System 2. As a way to live your life, however, continuous vigilance is not necessarily good, and it is certainly impractical" (2011, p. 28). Furthermore, "effort is required to maintain simultaneously in memory several ideas that require separate action" (p. 36) and "switching from one task to another is effortful, especially under time pressure" (p. 37).

The fast and slow thinking patterns of S1 and S2 apply to all areas of decision making including economics (BE), cybersecurity (BEC), and cyber operations (**BECO**). When cyberactors focus on absorbing tasks, they are oblivious to other important signals and commit biases that override their experience and training.

Incorporation of Biases in Economics, Cybersecurity, and CO

The integration of psychological findings of behavior and judgment into economics, i.e., the progression from SE to BE, required revisions of mainstream economic methods. According to Rabin, the difference between psychology and economics is that "while psychology investigates humans in all their richness, economics requires models that are not so rich as to retard the process of drawing out their economic implications" (1996, p. 2). Psychologists provide the breadth of information about the human psyche, and economists then use the filters of simplicity and tractability to select the psychological findings that enable them to build meaningful economic models.

Economic methods include methodological individualism, mathematical formalization of assumptions, logical analysis of the relationship between conclusions and assumptions, and empirical field testing. In SE, *methodological individualism* consists of two basic components: actors have well-defined preferences and they rationally maximize these preferences. BE revises these components by applying empirical evidence from psychology to the economic assumption-making to modify the nature of the preferences (Rabin, 1996, Section 2), demonstrate systematic errors that individuals commit when maximizing their utility functions (Rabin, 1996, Section 3), and describe scenarios where the very concept of people maximizing their preferences does not hold (Rabin, 1996, Section 4). Some cognition-based modifications are relatively easy to incorporate into economic models; other psychological findings raise awareness of the model shortcomings and improve economics on an *ad hoc* basis. Psychologists and experimental economists conduct controlled laboratory experiments to generate hypotheses, and economists test these hypotheses in uncontrolled field studies. Likewise, BECO is a hypothesis for integrating BE models into the CO Concepts of Operations (CONOPS) to be tested in field studies, as illustrated in Figure 5.

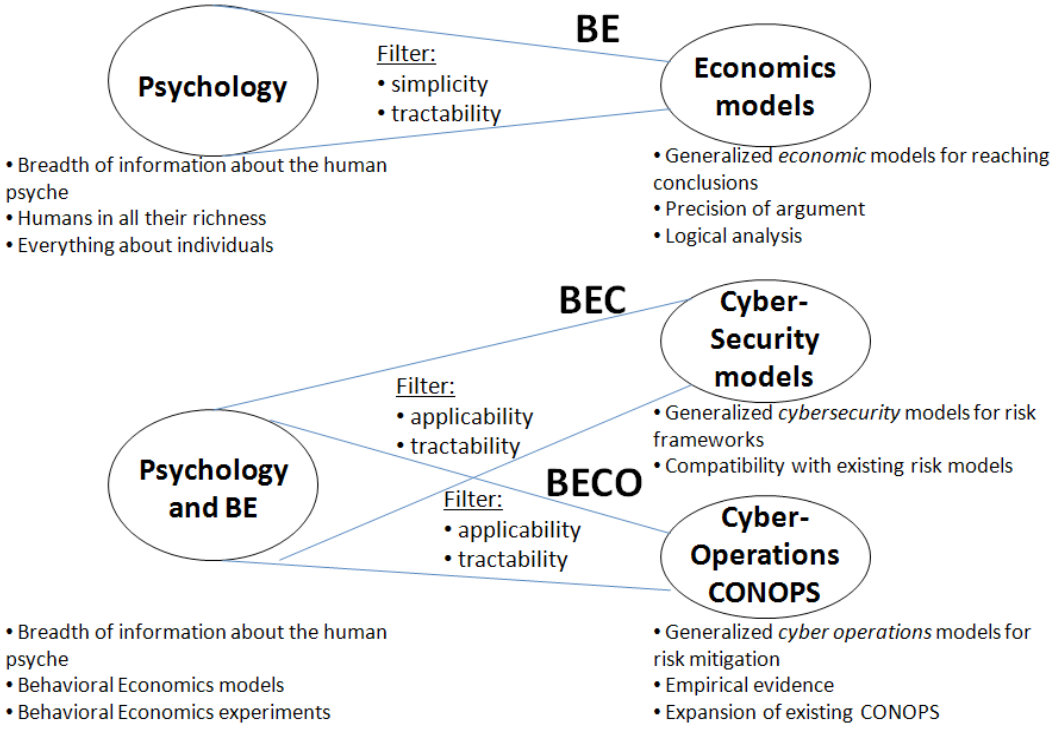


Figure 5. Relationships Between Psychology, BE, BEC, and BECO.

BECO will identify psychology and BE findings that could provide meaningful CONOPS enhancements. As with BE, some of these findings will be incorporated into CONOPS directly, while others will be used to raise awareness and improve the operations on an *ad hoc* basis.

BECO Solution and Innovation

BECO is a proposed framework for increasing the effectiveness of Cyberspace Operations, such as those of USCYBERCOM, by defining a risk management framework of the CO cognitive dimension. BECO identifies biases in the operational judgment and decision-making and seeks their mitigation on the friendly side and their exploitation on the adversary side. In this context, "the friendly side" refers to the United States and its allies, and "the adversary side" refers to states and non-state entities opposing the U.S. in cyberspace.

BECO Description

BEC model. BECO is an application of BEC to CO, where BEC is a framework for conducting BE-based cybersecurity risk management (Fineberg, 2014). BEC is defined in three *dimensions* of Cyberactors, Security Services, and Controls as depicted in Figure 6.

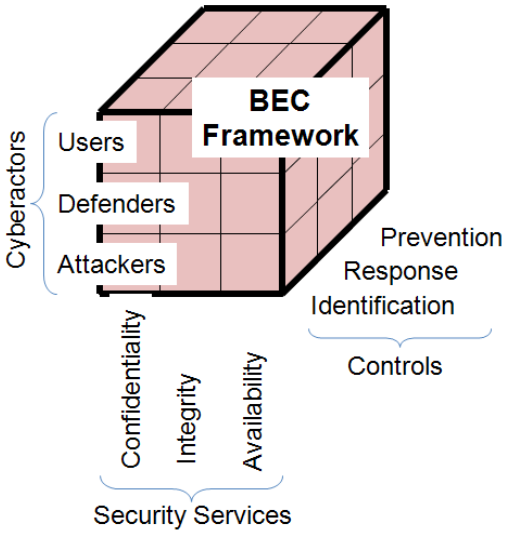


Figure 6. BEC framework.

Cyberactors are classes of individuals defined by their distinct cyber roles of Users, Defenders, and Attackers. Users are seeking functional capabilities of cyberspace, *Defenders* are protecting cyberspace, and *Attackers* are exploiting cyberspace. *Security Services* are classes of features that ensure proper cyberspace operation and include Confidentiality, Integrity, and Availability. *Confidentiality* is protection of the user information, *Integrity* is protection of cyber systems and data from unauthorized access and malicious manipulation, and *Availability* is the user's ability to use cyberspace systems and data. *Controls* are risk-management responses for upholding cybersecurity including Identification, Response, and Prevention. *Identification* uncovers significant cognitive biases that apply to various scenarios, *Response* mitigates biases on the friendly side and exploits biases on the adversary side, and *Prevention* encompasses research, training and other preparation.

The BEC cube can be used for comprehensive Risk Management and for selecting and controlling the greatest risks. In the Risk Assessment phase, cognitive vulnerabilities are represented by one or more squares on the Cyberactor-Security Services surface; and in the Risk Response phase, mitigation is selected along the Controls axis.

BECO model. BECO applies BEC to CO exemplified by the USCYBERCOM's mission. The principal distinctions between the two frameworks are their respective scopes and sets of actors. The scope of BEC is the general cybersecurity risk management, whereas the scope of BECO is risk management of the full-spectrum CO, as depicted in Figure 7. The BEC RMF is applied to each BECO actor, thus creating a five-dimensional analysis space of Cyberactors, Security Services, Controls, Planning Levels, and Lines of Operation.

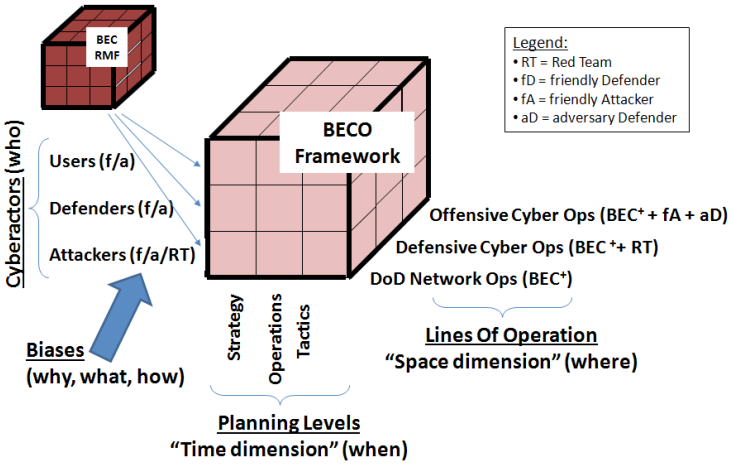


Figure 7. BECO framework.

A comprehensive scope of BECO is assured by its incorporation of a comprehensive set of questions "who, why, what, how, when, and where." "Who" are CO cyberactors, and "why, what, and how" represent actors' biases and actions. "When" is the time dimension, the timeframe of the strategic, operational, and tactical levels of the CO planning. "Where" is the space dimension, such as the USCYBECOM's Lines Of Operation (LOO) including DoD Network Operations (DNO), Defensive Cyber Operations (DCO), and Offensive Cyber Operations (OCO). DNO provides typical enterprise security within the

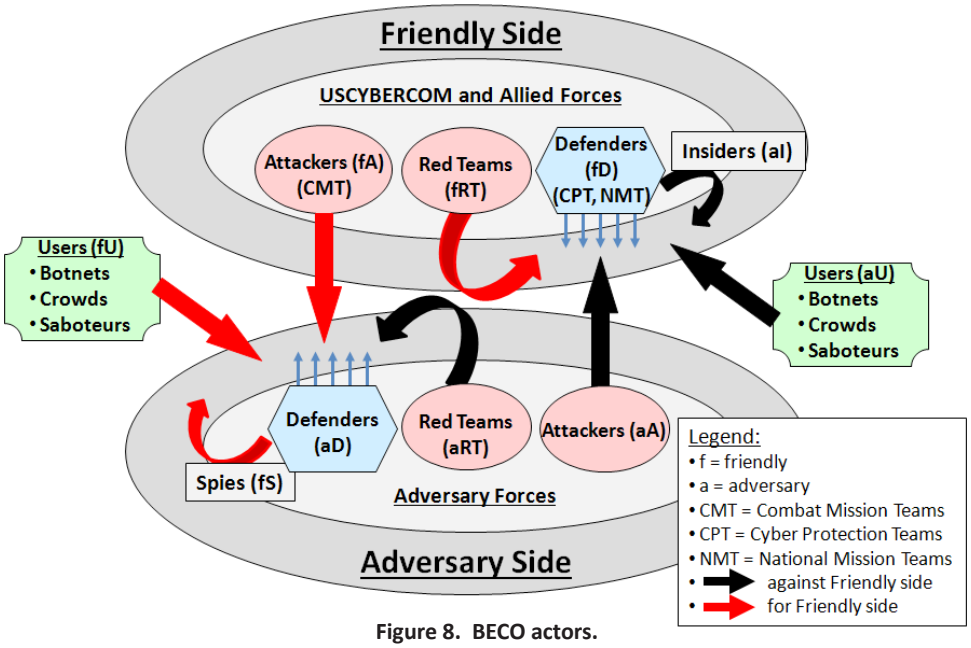


Figure 8. BECO actors.

defense perimeter, and its risk management corresponds to the original BEC. DCO extends DNO with the maneuver capability outside the perimeter and employs Red Teams. OCO engages in global military actions, in which USCYBERCOM's attackers are on the friendly side. In CO, the scope of actors expands beyond BEC's Users, Defenders, and Attackers by the considerations of the friendly and adversary sides as depicted in Figure 8, where the friendly-side USCYBECOM forces are described by Pellerin (2013b).

The *friendly* side includes Defenders (fD) such as USCYBERCOM's Cyber Protection Teams (CPT) and National Mission Teams (NMT), Attackers (fA) such as Combat Mission Teams (CMT), and Red Teams (fRT) testing the friendly defenses. On the *adversary* side, Attackers (aA) are regular BEC attackers and Defenders (aD) are BECO entities whose cognitive biases are exploited by fAs. Insiders (aI) are adversarial actors sabotaging the friendly side from inside the friendly defense perimeter; similarly, Spies (fS) are supporting the friendly side from inside the adversary defense perimeter. BECO Users include both adversary Users (aU) and friendly Users (fU) that may undermine the friendly and the adversary sides, respectively.

Examples of BECO Biases

Insider Biases and Mitigation

The BECO Black Swans are insiders (aI) that disclose information of vital importance or enable adversaries to penetrate the friendly defenses. Insiders may operate in various building blocks of the BECO framework and are particularly dangerous in *Strategy* and *Operations*. Like with all Black Swans (Taleb, 2010), their actions and motivations are analyzed *a posteriori*, and CO proceeds to fight the last war. The most dangerous insiders are turn-coat defenders. Their downfall is gradual, from minor infractions to full-blown national security violations. This pattern resembles *coherent arbitrariness* (Ariely, Loewenstein, & Prelec, 2000), a human propensity for mentally anchoring on arbitrarily selected initial conditions ("arbitrariness") and then making judgments systematically related to the initial selection ("coherence").

In Ariely's experiments, students were either paid or received payment for listening to his poetry recital, depending on whether a session was offered as an entertainment or a chore (2009). The students' initial decision to pay for Ariely's poetry was as arbitrary as that of Tom Sawyer's friends agreeing to pay for whitewashing his aunt's fence, but after that decision had been made, the amounts paid were coherently proportional to the duration of the experience. To fight the formation of undesirable patterns, Ariely urges decision makers to question their repeated behaviors and pay particular attention to the initial decisions in what is going to become a long stream of decisions. Likewise, in **BECO** it is important to identify and prevent decisions that may turn defenders into insiders and break harmful patterns as soon as they form.

Insider actions may also be forestalled by using a powerful psychological mechanism of *cognitive dissonance* described by Leon Festinger (1962) as "that if a person knows various things that are not psychologically consistent with one another, he will, in a variety of ways, try to make them more consistent" (p. 93). Ariely provides an example of how "doctors reason that if they are telling others about a drug, it must be good—and so their own beliefs change to correspond to their speech, and they start prescribing accordingly" (2012, p. 81). Furthermore, actions create preferences, because "decisions can be highly sensitive to situational factors, even when such factors are unrelated to the actual utility of that course of action" and individuals "rely not only on stable hedonic utilities but also on their memories of utility for their own past behaviors" (Ariely & Norton, 2008, p. 13). In **BECO**, cognitive dissonance can be used to enhance the loyalty of the cyber workforce by asking defenders to perform patriotic duties beyond their normal responsibilities and thereby develop a positive mindset. Negative attitudes of defenders must be curtailed before they deepen and lead to adversarial insider actions.

Hawkish Biases

Hawkish biases influence military strategy towards aggressive "hawkish" attitudes and downplay conciliatory, or "dovish," attitudes beyond common considerations of prudence. They affect attackers on the friendly and adversary sides in the BECO *strategy* block. Mitigation of these biases requires actions by attackers, defenders and Red Teams as discussed in this section. The name "hawkish biases" was introduced by Kahneman and Renshon (2009) who reviewed an extensive list of cognitive biases in the context of military and diplomatic actions and found that all of them were strongly directional towards aggression. The set of *positive illusions* includes "unrealistically positive views of one's abilities and character, the illusion of control, and unrealistic optimism" (p. 4). *Unrealistically positive views* lead people to consider themselves better decision makers and negotiators than they are. People experience the *illusion of control* when they exaggerate the impact of their actions on the outcomes, and under stress, they prefer strategies that they think would give them more control. *Unrealistic optimism* causes people to overestimate the odds of positive for them events, have more confidence in their predictions than the circumstances warrant, and discount the abilities and skills of their peer group. Political science studies and simulated conflicts have demonstrated that the positive-illusion biases cause leaders to have unrealistically positive views of the balance of military power, and many wars start because leaders on each side believe that they will win. Furthermore, during a conflict, negotiations stall because each

side thinks that it has a stronger hand and, therefore, it is less likely to make concessions. Positive illusions take place in **BECO** when each side overstates its attack capabilities against the other side's defenses. The attackers' illusion of control may be exploited by the opponent setting up deception systems.

The *Fundamental Attribution Error (FAE)* is a bias of explaining behaviors of others by exaggerating their intentions and discounting their circumstances. This bias persists even when people are aware of it. In conflicts, "beliefs in the hostile intentions of adversaries tend to be self-perpetuating—and … self-fulfilling" (p. 8), whereas the true reasons for hostile actions could be in response to the opponent's domestic politics or to one's own aggression. With the FAE, the hawkish behavior is prompted by attributing the opponent's moderate behavior to their situation and their hostile behavior to their disposition. The FAE affects **BECO** when each side assumes that the other side is preparing cyber attacks. Moreover, in cyberspace, the FAE may be exaggerated even further, because many exploits are invisible until they are launched. The players must be aware of the FAE and try to distinguish genuine attacks from random events before responding in kind. Considering the difficulty of attribution in cyberspace and the need for an almost instantaneous response, defenders must have effective diagnostic capabilities. When there is a possibility that an adversary may misperceive an attack, direct communications between decision makers are particularly important.

Loss aversion is a manifestation of people's greater sensitivity to losses than gains. Related biases are the *endowment effect* of overvaluing the items people already own in comparison to identical items that do not belong to them, and the *status-quo bias* of the preference for the existing situation even if a change would be more beneficial. Loss aversion negatively affects negotiations, because each side considers its concessions as greater losses than they are gains for the other side. In **BECO**, the endowment effect causes cyberactors to overestimate the merit of their strategies, processes, and technologies. Recommendations for significant changes must not only be justified logically but also address commanders' biases, and analyses of alternatives must be conducted by independent parties. *Confirmation bias* causes commanders to overvalue evidence supporting their beliefs that some types of cyber attacks are more likely, some adversaries are more dangerous, and some defenses are more effective. The BECO countermeasures should include independent reviews and stress tests by Red Teams.

Risk seeking in losses causes people facing a sure loss to take greater risks. In conflicts, the side anticipating a significant loss is prone to engage in a disastrous campaign that has a small chance of winning; instead of ignoring *sunk costs*, leaders escalate commitment . An *agency problem* compounds these effects, because leaders ("agents") are punished for losses and rewarded for gains even in situations where their constituency ("principals") would have preferred a loss to a foolish risk. In **BECO**, the players that consider themselves more vulnerable, e.g., non-state entities, may be attacking more aggressively to avoid a certain loss. Considering that in cyberspace the actual capabilities are concealed and perceptions are more potent than in physical realms, irrationally-motivated attacks are more likely. The agency problem is also evident in the botnet phenomena where the owners of infected computers ("adversary users," or aU) are "agents" who don't really suffer from the Distributed Denial of Service (DDoS) attacks that they precipitate.

Pseudo-certainty is a bias in multi-stage decision-making of choosing the certain outcome of the last stage while disregarding the probability of reaching that last stage. This situation frequently arises in international politics where decision makers focus on the certainties of the final stage and disregard the contingency of the final stage on preceding stages, which may strongly depend on the decision makers' choices. Thus, "actors under-emphasize the effect of their own actions" (p. 19). In **BECO**, this means that an actor may focus on its strength in a full-blown cyber conflict and disregard the statistical uncertainties of the preliminary actions leading to it. Field experiments may indicate that in BECO individual hawkish biases might benefit from a consolidated approach if any of them reinforce or diminish the influence of others.

BECO Mitigation

Biased decisions are frequently made when individuals are in a "hot" state, i.e., their reflexive thinking dominates their logical thinking (Ariely, 2009, pp. 120-121). This paper proposes a structured mitigation approach for preparing friendly-side cyberactors for potential hot states as depicted in Figure 9.

The mitigation framework covers a range of approaches starting with the hot state avoidance (1), proceeding to switching to a cold state in different parts of the process (2 and 3), and then moving to various approaches to the preparation to and management of the hot state itself (4 through 8). This framework formalizes recommendations found in discrete sources as follows:

1. Avoid some hot states all together (Ariely, 2009, pp. 130-131), because upon entering these states resistance to temptation becomes extremely difficult. A **BECO** example of such hot-state avoidance is blocking access to the Internet pornography sites.

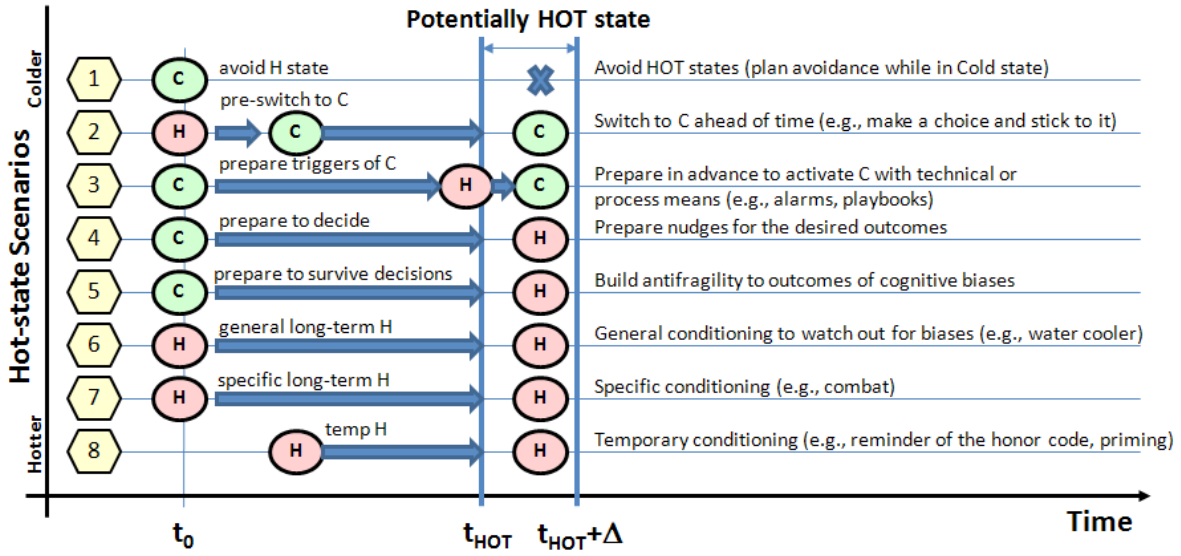


Figure 9. A structured approach to mitigating cognitive biases.

2. When a choice is difficult to make, because all options have approximately the same utility even as the details vary, choose one option and stick to it instead of prolonging the analysis and getting paralyzed by choice (Ariely, 2009, pp. 194-196). In **BECO**, this corresponds to choosing certain critical operational responses in advance.

3. Powerful technical and process controls must be defined to activate cyberactors' cold state at the point where they are likely to make critical mistakes, thus satisfying Kahneman's wish to have "a warning bell that rings loudly whenever we are about to make a serious error" (2011, p. 417). An existing example of such a control is an operating system that asks users to confirm that they want to delete a file. In **BECO**, Tactics, Techniques, and Procedures (TTPs) must be defined for anticipated critical decision points, forcing cyber warriors to invoke their System 2 thinking.

4. Nudges (Thaler & Sunstein, 2009) and defaults are used to suggest a preferred option without forcing it. In **BECO**, this approach is more applicable to cyber users who are free to choose than to warfighters who can be compelled to act in certain ways by their organizations.

5. Taleb (2010) urges to consider Black-Swan events not as exceptions that are explained *a posteriori*, but as a class of low-probability high-impact events that cannot be *individually* predicted. The best preparation for potential Black Swans is to cultivate antifragility (Taleb, 2012) that would protect an entity from a broad range of calamities. A current example of such preparation is Continuity Of Operations Planning (COOP), which Fineberg recommends enhancing with random stress testing (2012). Stress testing for developing antifragility should also be incorporated into a variety of **BECO** scenarios, e.g., cyber flag exercises (Alexander, 2012, p. 14).

6. Behavioral economists warn that the knowledge of cognitive biases does not prevent people from committing these biases. Kahneman admits that his intuitive thinking is just as prone to overconfidence and other System 1 manifestations as it was before he started studying these issues. However, he has improved his ability to *recognize situations* in which errors are likely, and once he recognizes them, to slow down and invoke System 2 (2011, p. 417). It is also easier to recognize errors of others than one's own, because "observers are less cognitively busy and more open to information than actions." Kahneman recommends having water-cooler discussions to take advantage of the group influences. **BECO** training should include developing the recognition of error-prone situations, and BECO CONOPS should include activities that activate group influences.

7. Conditioning for specific situations prepares people for taking the correct action when the situation arises, as for example, practiced by psychiatrists in the behavioral therapy of Exposure Response Prevention (ERP) for treating conditions such as panic. ERP practitioners select appropriate frequency, duration, rate of build-up, and escape prevention to achieve high levels of effectiveness. Likewise, certain combat situations require a single instantaneous decisive action. For physical combat, Grossman and Christensen recommend operand conditioning, i.e., realistic training until a warrior performs required actions automatically without thinking, because "whatever you rehearse is what you do under stress" (2007, p. 47). For example, practice shooting at moving targets shaped as human silhouettes has increased the front-line firing rate from 15 to 20 percent

in World War II to 90 percent during the Vietnam War. The **BECO** counterpart of such a conditioning is General Alexander's request for a single standard for taking action (2012, p. 14).

8. People can be prepared for decision making in a process called "priming," which is widely used in cognitive psychology experiments to affect the choices people make in a hot state. For example, Ariely (2009, p. 283) shows how reciting of the Ten Commandments prior to exams has resulted in significantly reduced student cheating. Likewise, in **BECO** cyber warfighters can be primed with the reminders of their honor code.

While cognitive biases have been extensively identified and thoroughly studied, their mitigation is challenging. A critical issue is that mitigation may work in laboratory experiments but not in real-life scenarios. Another problem is that any given mitigation may work in a short term but wear off with repetition. Nevertheless, the principal reason for identifying cognitive biases in BECO is the potential ability to develop effective responses. To facilitate research of mitigation, a full-scope cyber force such as USCYBERCOM can use its defense forces to test its attackers and use its attack forces to test its defenders as depicted in Figure 10.

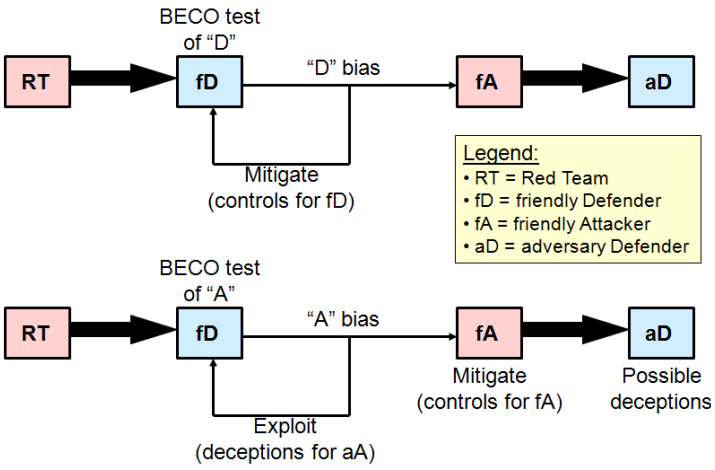


Figure 10. Bias testing architectures.

The top part of Figure 10 illustrates how Red Team (RT) probing of vulnerabilities of the friendly-side defenders can be used to strengthen friendly defenders (fD) and weaken adversary defenders (aD). For example, RTs may discover that defenders get accustomed to false alarms and start neglecting them. To mitigate this tendency with fDs, new TTPs will be implemented to vary the strength and appearance of the alarms using psychological techniques of irregular reinforcement. To exploit this tendency with aDs, friendly attackers (fA) will stage multiple false attacks before launching the actual attack.

The bottom part of Figure 10 illustrates how cognitive biases revealed by RTs can be used to strengthen friendly attackers (fA) and weaken adversary attackers (aA). For example, an attacker may be affected the paradox of choice, i.e., getting paralyzed with indecision when confronted with too many choices (Iyengar & Lepper, 2000). To exploit it, fDs can present to aAs many enticing choices. To mitigate it, fAs can be requested to follow strict decision making processes for selecting their targets and abandoning exploits upon reaching certain thresholds, thus eliminating the perils of choice.

These mitigating approaches and their details must be thoroughly researched and carefully implemented to provide the friendly side with tangible advantages in the cyber warfare. An important part of this research is the role of leaders and groups, who serve as psychological weapons (Grossman & Christensen, 2007, pp. 205-208) and, as most other organizations, "naturally think more slowly … and impose orderly procedures" (Kahneman, 2011, p. 418), thus mitigating the quirks of the individual human cognition.

Conclusions

This paper proposes a novel framework BECO of using the behavioral economics (BE) models of cognitive biases in judgment and decision making for hardening cyberspace operations (CO). BE adapts psychology research to economic models, thus creating more accurate representations of human interactions. BEC (Fineberg, 2014) uses BE discoveries to modify the risk management framework of cybersecurity by introducing a new class of vulnerabilities corresponding to persistent human biases. And now BECO applies the BEC framework to cyberspace operations by providing an overarching approach to the cognitive characteristics of the full spectrum of the CO actors and scenarios. Cyberspace operations are exemplified by the USCYBERCOM's mission, and cyberactors include attackers, defenders, and users on both the friendly and adversary sides. The paper reviews selected BE biases applicable to CO and offers a structured approach to the cognitive bias mitigation.

BECO provides an asymmetric advantage to cyber superpowers that have resources to research cognitive biases in their operations and implement effective controls. While non-state actors may obtain technologies developed by major states, they cannot replicate a unique operational environment of a cyber power. Furthermore, full scope forces, such as USCYBERCOM, can use their attack and defense capabilities to cross-test and strengthen the cognitive aspects of both. BECO goals are to define interdisciplinary research of cognition in cyberoperations, develop cyberoperations policies and strategies, and train cyber workforce.

About the Author

Victoria Fineberg is a Principal Information Assurance Engineer at the Defense Information Systems Agency (DISA). Victoria holds a Master of Science Degree in Mechanical Engineering (MSME) from the University of Illinois at Urbana-Champaign and a Master of Science Degree in Government Information Leadership with specialization in Cybersecurity (MS GIL-Cybersecurity) from the National Defense University's (NDU) iCollege. She is a licensed Professional Engineer and a Certified Information Systems Security Professional (CISSP). Prior to DISA Victoria worked at Bell Labs, Lucent Technologies. Her professional interests include cybersecurity, risk analysis, and the impact of cognitive biases on cyber operations.

References

Alexander, K. B. (2012). Statement before the Senate Committee on Armed Services. Retrieved from http://www.airforcemag.com/SiteCollectionDocuments/Reports/2012/March2012/Day28/032812alexander.pdf.

Ariely, D. (2009). *Predictably irrational: The hidden forces that shape our decisions*. Revised and expanded edition. New York, NY: Harper Perennial.

Ariely, D. (2012). *The (honest) truth about dishonesty: How we lie to everyone—Especially ourselves*. New York, NY: HarperCollins Publishers.

Ariely, D., Loewenstein, G., & Prelec, D. (2000). Coherent arbitrariness: Duration-sensitive pricing of hedonic stimuli around an arbitrary anchor. SSRN. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=243109.

Ariely, D. & Norton, M. I. (2008). How actions create – not just reveal – preferences. *Trends in Cognitive Sciences*, 12 (1), 13-16.

Festinger, L. (1962). Cognitive dissonance. *Scientific American*, 207(4), 93-107.

Fineberg, V. (2012). COOP hardening against Black Swans. *The Business Continuity and Resiliency Journal*, 1(3), 14-24.

Fineberg, V. (2014). BEC: Applying behavioral economics to harden cyberspace. *Journal of Cybersecurity and Information Systems*, 2(1), 27-33. Retrieved from https://www.csiac.org/journal_article/bec-applying-behavioral-economics-harden-cyberspace#.U6nby7H5eZk.

Grossman, D. & Christensen, L. W. (2007). *On combat: The psychology and physiology of deadly conflict in war and in peace*. 2nd Edition. PPCT Research Publications.

Holton, J. W. (2011). *The Pashtun behavior economy: An analysis of decision making in tribal society*. Master's Thesis. Naval Postgraduate School. Monterey, CA. Retrieved from http://edocs.nps.edu/npspubs/scholarly/theses/2011/June/11Jun_Holton.pdf.

Iyengar, S. S. & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6), 995-1006. Retrieved from http://www.columbia.edu/~ss957/articles/Choice_is_Demotivating.pdf.

JP 3-13. (2012). *Information operations*. Joint Publication 3-13. Retrieved from http://www.dtic.mil/doctrine/new_pubs/jp3_13.pdf.

Kahneman, D. (2006). [Video File]. History and rationality lecture series. Hebrew University. Retrieved from http://www.youtube.com/watch?v=3CWm3i74mHI.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D. (2013). [Video File]. Annual Hans Maeder lecture with Nobel Prize-winning psychologist Daniel Kahneman. The New School. Retrieved from http://www.youtube.com/watch?v=l91ahHR5-i0&list=PLUWrLGgGJAm9pm4ANtiGk4VVflf45Hz0P&index=7.

Kahneman, D. & Renshon, J. (2009). Hawkish biases. Expanded version of an essay that appeared in *American Foreign Policy and the Politics of Fear: Threat Inflation Since 9/11*. New York, NY: Routledge Press, 79-96. Retrieved from http://www.princeton.edu/~kahneman/docs/Publications/Hawkish%20Biases.pdf

Mackay, A. & Tatham S. (2011). *Behavioural conflict: Why understanding people and their motives will prove decisive in future conflict*. Saffron Walden, Essex, UK: Military Studies Press.

NIST 800-39. (2011). Managing information security risk: Organization, mission, and information system view. *NIST Special Publication* 800-39. Gaithersburg, MD: Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology. Retrieved from http://csrc.nist.gov/publications/nistpubs/800-39/SP800-39-final.pdf.

Pellerin, C. (2013a). Cyber Command adapts to understand cyber battlespace. U.S. Department of Defense. Retrieved from http://www.defense.gov/news/newsarticle.aspx?id=119470.

Pellerin, C. (2013b). DOD readies elements crucial to Cyber Operations. U.S. Department of Defense. Retrieved from http://www.defense.gov/news/newsarticle.aspx?id=120381.

Rabin, M. (1996). Psychology and Economics. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.42.9558&rep=rep1&type=pdf.

Stavridis, J. G. & Parker, E. C. III. (2012). Sailing the cyber sea. JFQ, 65(2), 61-67.

Taleb, N. N. (2010). *The Black Swan: The impact of the highly improbable*. New York, NY: Random House.

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder*. New York, NY: Random House.

Thaler, R. H. & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. London, England: Penguin Books.

U.S. Cyber Command. (2013). United States Strategic Command factsheet: U.S. Cyber Command. Retrieved from http://www.stratcom.mil/factsheets/Cyber_Command/.

Article Submission Policy

The CSIAC Journal is a quarterly journal focusing on scientific and technical research & development, methods and processes, policies and standards, security, reliability, quality, and lessons learned case histories. CSIAC accepts articles submitted by the professional community for consideration. CSIAC will review articles and assist candidate authors in creating the final draft if the article is selected for publication. However, we cannot guarantee publication within a fixed time frame.

Note that CSIAC does not pay for articles published.

AUTHOR BIOS AND CONTACT INFORMATION

When you submit your article to CSIAC, you also need to submit a brief bio, which is printed at the end of your article. Additionally, CSIAC requests that you provide contact information (email and/or phone and/or web address), which is also published with your article so that readers may follow up with you. You also need to send CSIAC your preferred mailing address for receipt of the Journal in printed format. All authors receive 5 complementary copies of the Journal issue in which their article appears and are automatically registered to receive future issues of Journal

COPYRIGHT:

Submittal of an original and previously unpublished article constitutes a transfer of ownership for First Publication Rights for a period of ninety days following publication. After this ninety day period full copyright ownership returns to the author. CSIAC always grants permission to reprint or distribute the article once published, as long as attribution is provided for CSIAC as the publisher and the Journal issue in which the article appeared is cited. The primary reason for CSIAC holding the copyright is to insure that the same article is not published simultaneously in other trade journals. The Journal enjoys a reputation of outstanding quality and value. We distribute the Journal to more than 30,000 registered CSIAC patrons free of charge and we publish it on our website where thousands of viewers read the articles each week.

FOR INVITED AUTHORS:

CSIAC typically allocates the author one month to prepare an initial draft. Then, upon receipt of an initial draft, CSIAC reviews the article and works with the author to create a final draft; we allow 2 to 3 weeks for this process. CSIAC expects to have a final draft of the article ready for publication no later than 2 months after the author accepts our initial invitation.

PREFERRED FORMATS:

- Articles must be submitted electronically.
- MS-Word, or Open Office equivalent (something that can be edited by CSIAC)

SIZE GUIDELINES:

- Minimum of 1,500 – 2,000 words (3-4 typed pages using Times New Roman 12 pt font) Maximum of 12 pages
- Authors have latitude to adjust the size as necessary to communicate their message

IMAGES:

- Graphics and Images are encouraged.
- Print quality, 300 or better DPI. JPG or PNG format preferred

Note: Please embed the graphic images into your article to clarify where they should go but send the graphics as separate files when you submit the final draft of the article. This makes it easier should the graphics need to be changed or resized.

CONTACT INFORMATION:

CSIAC
100 Seymour Road Suite C102
Utica, NY 13502
Phone: (800) 214-7921
Fax: 315-351-4209

John Dingman, Managing Editor
Email: jdingman@quanterion.com

Michael Weir, CSIAC Director
Email: mweir@quanterion.com

CSIAC is a DoD sponsored Information Analysis Center (IAC), administratively managed by the Defense Technical Information Center (DTIC), technically managed by the Air Force Research Laboratory (AFRL) in Rome, NY and operated by Quanterion Solutions Incorporated, Utica, NY.

ABOUT THE JOURNAL OF CYBER SECURITY AND INFORMATION SYSTEMS

CSIAC JOURNAL EDITORIAL BOARD

Distribution Statement
Unclassified and Unlimited

CSIAC
100 Seymour Road
Utica, NY 13502-1348
Phone: 800-214-7921 • **Fax:** 315-732-3261
E-mail: info@csiac.org
URL: https://www.csiac.org/

ABOUT THIS PUBLICATION

The Journal of Cyber Security and Information Systems is published quarterly by the Cyber Security and Information Systems Information Analysis Center (CSIAC). The CSIAC is a DoD sponsored Information Analysis Center (IAC), administratively managed by the Defense Technical Information Center (DTIC). The CSIAC is technically managed by Air Force Research Laboratory in Rome, NY and operated by Quanterion Solutions Incorporated in Utica, NY.

COVER DESIGN

Shelley Howard
Graphic Designer
Quanterion Solutions, CSIAC

ARTICLE REPRODUCTION

Images and information presented in these articles may be reproduced as long as the following message is noted:

"This article was originally published in the Journal of Cyber Security and Information Systems Vol.II, No 2"

In addition to this print message, we ask that you notify CSIAC regarding any document that references any article appearing in the *CSIAC Journal.*

Requests for copies of the referenced journal may be submitted to the following address:

Cyber Security and Information Systems
100 Seymour Road
Utica, NY 13502-1348

Phone: 800-214-7921
Fax: 315-732-3261
E-mail: info@csiac.org

An archive of past newsletters is available at **https://journal.csiac.org.**

Cyber Security and Information Systems
Information Analysis Center
100 Seymour Road
Suite C-102
Utica, NY 13502

Return Service Requested

PRSRT STD
U.S. Postage
P A I D
Permit #566
UTICA, NY

Journal of Cyber Security and Information Systems – Volume II Number 2
Games People Play - Behavior and Security

— IN THIS ISSUE —