Cybersecurity Strategies: The QuERIES Methodology

QuERIES offers a novel multidisciplinary approach to quantifying risk associated with security technologies resulting in investment-efficient cybersecurity strategies.

Lawrence Carin
Duke University

George Cybenko
Dartmouth College

Jeff Hughes
Air Force Research Laboratory

Organizations in both the private and public sectors have struggled to determine the appropriate investments to make for protecting their critical intellectual property. As a result, they have typically implemented cybersecurity investment strategies without useful guidance from a rigorous, quantitative risk-assessment and -mitigation methodology. Simple questions such as how much to invest, which security measures will have the most impact, and gauging the level of improvement in security currently prove difficult to answer.[1]

Designed to answer these questions, Quantitative Evaluation of Risk for Investment Efficient Strategies (QuERIES) offers a novel computational approach to quantitative cybersecurity risk assessment. We based this approach on rigorous quantitative techniques drawn from computer science, game theory, control theory, and economics. Preliminary experiments have corroborated the QuERIES methodology, suggesting that it provides a broadly applicable alternative to red teaming (which involves attackers who have little or no knowledge of a system's internal protection), black-hat analysis (which involves attackers who have access to design details of the internal protection), and other decision-support methodologies previously tried in cybersecurity-related risk assessment.

To date, QuERIES has focused on the problem of protecting critical US Department of Defense (DoD) intellectual property, in which the loss of one IP copy is catastrophic, as opposed to consumer IP, in which the loss of multiple copies can be tolerated if sufficient revenue can be maintained. Weapons systems designs, chip designs, complex computer software, and databases containing personal and financial information are examples of the former. Digital music, video, consumer-grade software, and electronic books are examples of the latter. Cybersecurity experts can apply QuERIES

Published by the IEEE Computer Society

0018-9162/08/$25.00 © 2008 IEEE

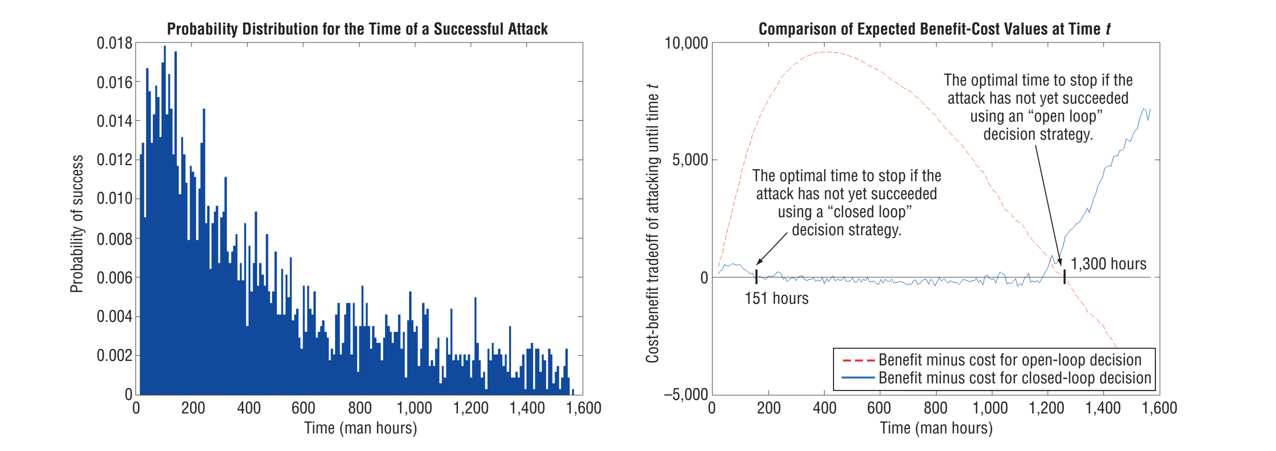

Figure 1. These plots show (left) the probability distribution for the time to achieve a successful attack and (right) the associated cost-benefit analysis.

to other attack or protect scenarios by appropriately changing the underlying economic model.

HOW IT WORKS

To illustrate the QuERIES methodology and how developers can apply it in a given software protection context, consider the challenge of assessing the strength of protections applied to a particular software asset. The protections are meant to prevent reverse-engineering attacks in which an adversary seeks to obtain critical IP from the software. The QuERIES methodology in this case involves the following elements.

Model the security strategy

This element develops an attack/protect economic model cast in game-theoretic terms. Parameters in this model represent objective quantities such as the economic value of the IP (the protected software asset) to the IP owner; what it would cost an adversary to develop the IP; and the cost of obtaining the IP through other possible means. Another critical ingredient of the model is the protection map (a detailed security plan) of the specific protections applied to the IP asset.

Model the attacks

This element uses the protection map and knowledge of reverse-engineering methodologies to build an attack graph represented as a Partially Observable Markov Decision Process (POMDP).[2]

Quantify both models

This element quantifies parameters used in both models by performing a controlled red-team attack against the protected IP, then using another red- or black-hat team to conduct an information market[3] for estimating the POMDP's parameters. It then computes the POMDP's optimal policies and uses those policies in the attack/protect economic model. Once the system has evaluated both models, synthesizing multiple derived quantities relevant to risk assessment becomes possible.

For example, given a class of adversaries, the left plot in Figure 1 shows one such derived quantity, namely the probability distribution of the time in man-hours required to successfully reverse-engineer protected software. We call this distribution the Probability of Reverse Engineering or P_R. This distribution assumes that the attacker does not have an a priori model of the protection scheme. The attacker therefore learns the protection scheme through trial and error. The probability distribution is a sampling of multiple independent attacks under this assumption.

The right plot in Figure 1 shows the results of two different analyses an attacker could use to decide when to stop an attack, namely open- and closed-loop decision algorithms. The results of different analyses can be quite different. Using the closed-loop decision algorithm, if the attacker has not succeeded after about 151 hours, the optimal decision is to stop the attack because it has reached the tail of the distribution. The probability of defeating the protections using that strategy is about 0.25, and the maximum cost (defined as the expected cost of a successful attack before time $t < 151$ plus the expected cost of failure at time $t = 151$) is about $7,895 compared with the assumed $30,000 initial value of the IP in this example.

The right plot also shows the difference between the expected costs and expected benefits (expected IP value over time) of conducting an attack up to the specified time plotted on the horizontal axis. This is an "open loop" analysis in that it does not factor in the attacker's stopping the attack even after passing the "fat" part of the probability distribution and thereby working with a diminished likelihood of returns. That analysis suggests that the attack's cost exceeds its benefits after 1,300 hours. In this example, we model cost by a constant $60 per man-hour of effort.

The probability distribution P_R that QuERIES obtains can be the basis for different kinds of analyses. Because

		You (Y)	
		No IP protection	**IP protection**
Adversary takes no action		Y: C_{IP} A: 0	Y: C_{IP}- C_P A: 0
Adversary develops IP		Y: C_{IP} - C_P A: C_{IP}- C_D	Y: C_{IP} - C_P A: - C_{IP}- C_D
Adversary steals IP with Prob = P_S or P_R and Cost = C_S or C_A	**Failure**	Y: C_{IP} A: C_S Prob = 1 - P_S	Y: C_{IP} - C_P A: - C_R Prob = 1 - P_R
	Success	Y: C_{IP} A: C_{IP} - C_S Prob = P_S	Y: C_{IP} - C_P A: C_{IP} - C_R Prob = P_R
	Expected	Y: C_{IP} A: $P_S C_{IP}$ - C_S	Y: C_{IP} - C_P A: $P_R C_{IP}$ - C_R

Figure 2. In this example, the QuERIES economic model is based on a simple game-theoretic formulation. In the game, the IP owner can protect or not protect and the adversary can develop the IP ab initio or attempt to steal or reverse-engineer it. Although the case in which the adversary chooses to do nothing is listed, the definition of critical IP is that the adversary will try to obtain the IP.

QuERIES is agnostic about how a decision maker actually uses P_R, we introduced the above derivative analyses solely to illustrate the fundamental role that it plays.

We believe that a major innovation of QuERIES is a methodology for estimating the fundamental distribution P_R. Traditional approaches for evaluating the strength of cybersecurity technologies have not been able to effectively produce the probability distribution of the time to defeat a protection.[4] For example, formal methods—logical analyses of a design—can only verify that a design has certain desirable properties; they are silent on an actual implementation's properties and its deployment in a complex operational environment. Red-team attacks as traditionally conducted result in a very sparse sampling of the distribution P_R, often producing only a single costly sample—namely that the attack took so much time, so many resources, and used a certain approach. Black-hat analyses typically suggest multiple possible attack paths and the associated tools required, with gross estimates of attack times and costs.[5]

QUERIES METHODOLOGY

QuERIES methodology users must first identify their critical IP assets and the threats against them through analysis of their various missions and strategic plans. We use a relatively objective measure of such an asset's value—the cost to develop it. Those costs usually can be estimated reliably using programmatic information, although in many cases the development of advanced systems leverages a broad technology base that might already have been expensed elsewhere. Our notation for the owner's cost of developing the IP is C_{IP}.

By definition, an adversary values critical IP at C_{IP} as well, but the development cost to an adversary, denoted by C_D, could be smaller if generally available enabling technology has made it more economical to develop today as opposed to in the past.

Hence the first step of the QuERIES method identifies the following:

- C_{IP}: the value of the IP to the asset owner and adversary;
- C_P: the cost of protecting the IP, per unit, together with a possible amortization of the protection technology's cost over the number of units to be protected;
- C_D: the cost to the adversary of developing the IP from inception;
- P_S: the probability of stealing the unprotected IP, based, for example, on historical data for similar IP; and
- C_S: the cost of stealing the unprotected IP, based on historical data for similar IP.

The developer could estimate these quantities for different adversaries who have different technology bases from which to recreate the IP and different capabilities for stealing the unprotected IP.

Constructing the Attack/Protect economic model

The QuERIES attack/protect economic model is a game with two players: the protector and the attacker. Game theory is a mature discipline originally developed to support strategic decision making, but now widely used for business and economic applications as well.[6]

As Figure 2 shows, the two basic game moves available to the protector are protect or do not protect critical IP. Different protection technologies are possible for a given IP, so in practice the protector has several possible moves, one for each protection type considered. In this example, we model three possible attacker moves: No Action, Develops IP, and Steals IP.

By the definition of critical IP, the adversary will try to either develop or steal the IP. For each combination of moves by the protector and attacker, we write down an expression for the resulting loss or gain in the corresponding game table cell. When an adversary attempts to steal or reverse-engineer critical IP, the probability of success is P_S and P_R, respectively.

Game accounts analysis

The QuERIES game analysis accounts for several player objectives. The IP asset owner wants to maximize its minimal advantage over the adversary. The advantage is, by definition, the difference between the owner's value and the adversary's value because the owner already has the IP. The adversary wants to maximize its value without factoring in the owner's value because it wants the IP, and the owner already has it. The IP asset owner moves first because it gets to decide whether to deploy protections or not when fielding the critical IP.

If the IP owner does not protect the IP, the adversary might attempt to steal it. For example, the adversary should be motivated to steal the IP if the probability of successfully stealing it, P_S, approaches 1, while the cost of stealing it, C_S, is small, so that $P_S * C_{IP} - C_S > C_{IP} - C_D$, where the cost to the adversary of developing the IP, C_D, typically approximates the C_{IP}.

On the other hand, if the IP owner does protect the IP, the adversary will attempt to reverse-engineer the protected IP when $P_R * C_{IP} - C_R > C_{IP} - C_D$ and will develop the IP from its inception otherwise. The IP owner must take into account the cost of protecting the IP to avoid having a Pyrrhic victory by making the protection cost more expensive than the cost of obtaining the IP for the adversary.

Possible actions could include executing the code in a debugging environment and modifying the executable in various ways.

Consequently, the IP owner should protect the critical IP asset if both

$$C_D - C_p > (1 - P_S) * C_{IP} + C_S \quad \text{and}$$
$$(1 - P_R) * C_{IP} + C_R - C_p > (1 - P_S) * C_{IP} + C_S.$$

These two inequalities have similar interpretations. The right sides of both inequalities are the same and represent the relative advantage the owner has over the adversary when the IP owner does not protect the IP asset, namely the difference between the owner's value, C_{IP}, and the expected value if the adversary chooses to steal the IP, $P_S * C_{IP} - C_S$, which we have seen that an adversary will do in realistic situations. The left sides of the inequalities are the IP owner's relative advantages if the owner chooses to protect in those cases when the adversary develops or reverse-engineers the IP, respectively.

This economic analysis requires several quantities: C_{IP}, C_p, C_D, P_S, and C_S, which can be estimated from available empirical data. The quantities that cannot readily be estimated from historical data are C_R and P_R. The next few steps of the QuERIES methodology involve estimating these quantities and the subsequent derivatives.

CONSTRUCTING THE POMDP

The methodology obtains effective estimates of the remaining quantities used in the economic model, namely the probability, P_R, and cost, C_R, of defeating the protected IP asset.

The QuERIES approach to estimating the probabilities and costs of successful attacks begins by first formulating the possible attacks based on the given protections in terms of an attack graph that represents the adversary's multistage decision processes and states of knowledge. Researchers have used attack graphs effectively in computer security studies.[7] QuERIES advances this concept in several fundamental ways. Its most fundamental contribution to cybersecurity attack modeling might well be the introduction to the field of POMDPs[2] and information markets.

A QuERIES POMDP is defined by a finite set of states, one of which the agent occupies at any given time. In the QuERIES attack modeling framework, the underlying states represent the attacker's progress toward defeating the IP protections, of which there are typically several. The start state corresponds to none of the protections being defeated, and the end state corresponds to successful reverse engineering of the IP, which might or might not require defeating all the protections used.

An attacker takes actions to defeat the protections that a protector has applied to the critical IP. Possible actions could include executing the code in a debugging environment and modifying the executable in various ways. One consequence of such actions is that the attacker moves from state to state, possibly remaining in the same state. Given an attacker's action, a Markov process models the probability of transitioning from one state to another, specifically a finite Markov chain in this case. That is, for every action, a Markov chain labeled by that action specifies the state transition probabilities resulting from the action. Another consequence, which depends on the action taken and the current state, is that some cost to the attacker is incurred. Such a modeling formalism is called a Markov decision process (MDP).[2]

While MDPs provide the basic modeling formalism, they are not quite enough. At any given time, an attacker typically does not know what state he is in because he cannot observe the states directly. In the QuERIES context, attackers might not know what protections have been deployed or defeated, and they cannot be certain about what penalties might have been introduced through their previous actions. POMDPs can model MDPs in which states are not directly observable.

This QuERIES methodology step outputs a POMDP structure that encapsulates the procedural structure of possible attacks against protected IP. By structure, we mean the collection of possible actions, states, and observables that can arise in an attack. The POMDP's complete specification must also include the various state transition probabilities, costs associated with taking certain actions in certain states, and the probabilities of making certain observations conditioned on being in a true state not directly observable by the attacker. Controlled red-team attacks combined with subsequent information or decision markets can estimate these quantities.

INFORMATION MARKETS

At first, it would seem that the problem of generating P_R and C_R has been made more difficult because a POMDP involves a large number of probabilities and costs corresponding to each state in the underlying MDP. QuERIES provides a fundamental insight in that the protector can obtain the underlying POMDP parameters from red teams but not in the traditional way. Using information or decision markets, estimating with high accuracy the underlying POMDP parameters becomes possible, as does computing optimal policies for the POMDP. Simulations using those policies estimate the probability distribution of P_R. The cost, C_R, is derived from P_R, assuming a constant labor rate per hour.

QuERIES uses information market mechanisms to estimate the QuERIES attack model POMDP parameters.

Market mechanisms

Participants can use information markets to interact with each other using simple exchanges of structured information. The outcome of an information market is a collective—not a consensus—estimate of a quantity. Information markets have recently received a great deal of attention in the popular and technical literature to forecast sales, market trends, and complex system behaviors.[8]

Examples of information markets include traditional financial markets like stock and commodities exchanges in which participants buy and sell shares, options, and various derivatives. The only information that participants effectively exchange in those markets consists of the prices at which they are willing to buy and sell various instruments and in what quantities. Those prices are the markets' estimate of the instrument's value, such as the valuation of a company or the future prices of commodities such as oil, orange juice, or lumber products. Pari-mutuel betting, such as at horse racetracks, also involves the exchange of information through the tote board odds for a race. The odds have a natural, intrinsic interpretation as probabilities for how the different horses will perform.

Estimation tools

The information markets' effectiveness, if properly constituted, is no longer controversial. Researchers have discovered mechanisms for effectively estimating probability distributions over large combinatorial spaces in which the number of individual elements can be extremely large.[9]

QuERIES uses information market mechanisms to estimate the QuERIES attack model POMDP parameters. Developers create a market red team whose purpose is not simply to defeat a protection. Instead, the team receives the protected IP and then participates in several information markets, each of which estimates different probability distributions relevant to the POMDP. The team can inspect the protected IP, attempt to defeat it in various ways, and otherwise educate itself on the protections' details. The market is real. There are financial incentives for making correct predictions of probabilities, just as in pari-mutuel horse race betting.

A different red team then actually conducts a traditional red-team exercise to identify what is "correct" to determine payouts. The fundamental outcomes of the market are the estimates of probability distributions and costs, not the specific traditional red-team exercise outcome. Using horse racing as an analogy, running a race once generates only a single sample from the probability distribution over the horses about which one will win. The information market premise asserts that if the horse race could be run repeatedly, the number of times that individual horses would win would converge to the number predicted by the odds, which are estimated by betting—the market in this case.

QuERIES and probability

QuERIES uses information markets to estimate the probability distributions underlying the POMDP and uses the final traditional red-team attack to "run the horse race" and determine payouts. Having an objective outcome and real incentives to perform well is considered critical to the information market concept's effectiveness.

We have conducted several red-team information markets to validate this approach as it is applied by QuERIES in the cybersecurity risk assessment problem domain. The various quantitative results we have shown were obtained by the QuERIES methodology using actual red-team information markets.

COMPUTE THE POMDP'S OPTIMAL POLICIES

Once the QuERIES information market has produced estimates of the POMDP's probabilities and costs, it can use standard techniques for finding optimal policies of POMDPs to determine the optimal action to take in each state to minimize a cost objective.[2] A common objective to optimize is cost, which, in the case of protecting critical IP, can be measured in time to defeat the protections.

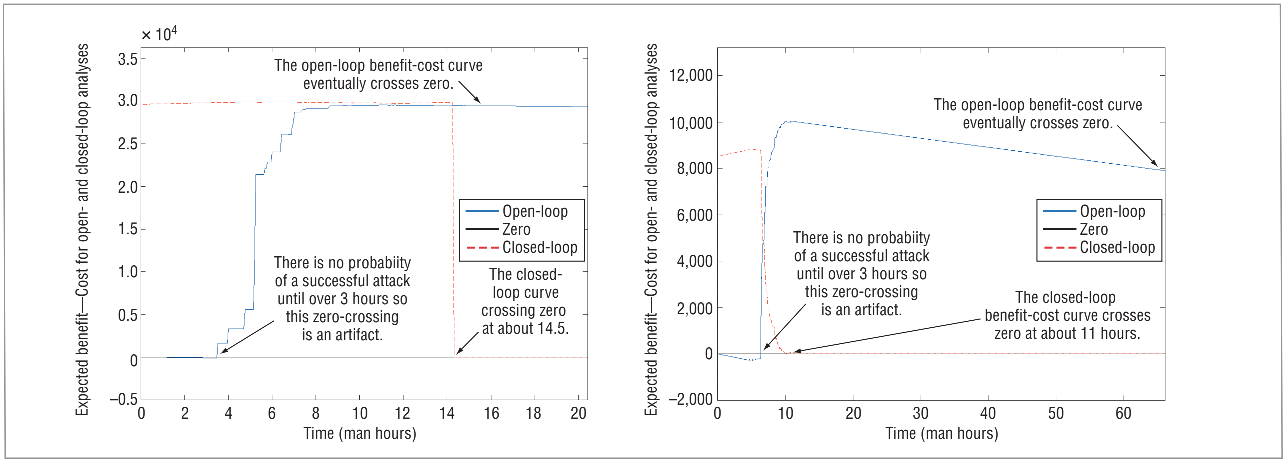

Figure 3. QuERIES analysis comparing two protections. The two plots compare the open-loop and closed-loop benefit minus cost curves and associated stopping times for (a) three protections and (b) four protections.

The optimal policy prescribes the action to take in each state of the POMDP to minimize the expected cost/time. A POMDP simulation that produces multiple runs using the optimal policy in every step will generate an empirical probability distribution of times to defeat the protections assuming that the attacker has knowledge of the optimal attack policy. By contrast, Figure 1 shows the probability distribution of successful attack times for an attacker who does not know the underlying structure (attack graph based on the protection map) or the POMDP's optimal policy before starting the attack.

By sampling from the policy space, it is possible to generate empirical distributions corresponding to suboptimal policies that represent less skilled or capable adversaries. For example, it is possible to randomly sample the second or third best policies to gain insight into different threat classes. By the same token, solely sampling from the optimal policy produces a distribution for P_R that corresponds to the most skilled attacker relative to the red-team skill level specified for the information markets.

Evaluate the Attack/Protect model

The previous steps produce optimal and suboptimal attack policies that the QuERIES methodology can use to generate a variety of probability distributions for the time and cost of successfully defeating the applied protections. Using those probability distributions, there are open-loop and closed-loop strategies for executing attacks that assume an attacker has some knowledge of P_R. The open-loop strategy does not take into account that an attack has not succeeded as it progresses. The closed-loop strategy does. The algorithmic basis for computing this strategy is similar to pricing a US-style financial option.[10] Further, if an attacker does not know P_R, the attacker's strategy will be based on some historical experience that entity might have.

Once QuERIES has estimated a distribution for P_R, it can derive corresponding estimates for C_R. The following example shows the kind of analysis possible using QuERIES to answer questions concerning the improvements resulting from additional protection investment, without estimating C_{IP}, C_P, C_D, P_S, or C_S, which lie outside QuERIES's scope.

Using QuERIES to compare protections

This QuERIES example can help a protector answer fundamental questions such as how well its IP is protected, how to determine the right level of investment in its protection, and determining the cost-benefit analysis for adding more protections. To illustrate this, we performed a worst-case analysis for two different protections.

The analysis in Figure 1 assumes that the attacker does not know the underlying POMDP attack graph model or parameters and so must expend effort exploring the protection and attack space. Figure 1 also shows the resulting probability distribution of time to defeat.

Given the POMDP model structure and parameter values, we can compute the probability distribution of the time to defeat, assuming that the attacker knows the optimal attack policy as specified by the POMDP attack graph structure and parameters. With this additional information, the time to defeat the protection is two orders of magnitude smaller when compared with Figure 1's results. We can repeat this analysis after adding another protection layer. With four protections, successful attacks appeared with only a probability of 0.35, which translates into a 0.65 probability that an attack will be unsuccessful in any reasonable time.

The two plots in Figure 3 compare the open-loop and closed-loop benefit minus cost curves and associated stopping times for three protections (left) and four protections (right). Even though successful attacks can occur in both cases, we can discover the value of the added protection in concrete, explicit, and quantitative terms.

To explain this analysis, note that with three protections, the optimal closed-loop policy for an attacker mandates attacking for about 14.5 hours, by which time the probability of a successful attack is 1. Moreover, the expected time to achieve a successful attack is 5.43 hours, which results in an expected cost of 5.43 × 60 = $326 for a successful attack and a benefit of about $30,000 (the decrease in IP value is negligible).

With four protections, the optimal policy attacks for only about 11 hours, with a success probability of only 0.35. The expected number of hours prior to the 11-hour stopping time is 2.5. With a 0.65 probability, the attacker will expend those 11 hours but fail. Therefore the expected cost (at $60 per hour) is 2.5 × 60 + 11 × 0.65 × 60 = $579, and the expected benefit is only 0.35 × 30,000 = $10,500.

There is only a 0.35 probability of successfully attacking the four protections because we use a bounded, fixed-resource model for the attacker. That is, the adversary's representational capacity and computing power are the same for three and four protections. With four protections, those resources are not sufficient to conduct successful attacks more than 35 percent of the time an attack is attempted. This resource limitation arises intrinsically when numerically computing optimal policies for POMDPs. Briefly, we use the same-sized approximation to represent the probability distributions underlying both three- and four-protection POMDP models.

In any case, as a result we can conclude that with three protections the attacker's expected gain is $30,000 – $326 = $29,674; with four protections, the attacker's gain is $10,500 – $579 = $9,921. Therefore, the added "insurance" of the fourth protection safeguards an additional $29,674 – 9,921 = $19,953 of the IP value.

Researchers can use the QuERIES methodology to rigorously determine, for the first time, appropriate investment levels and strategies for the protection of intellectual property in complex systems. As a result, it can have a significant and immediate impact on the protection of critical IP, including weapons systems and chip designs, complex computer software, and databases containing personal and financial information.

We have performed initial testing of QuERIES in small-scale, realistic scenarios, with results that suggest the methodology can significantly improve risk assessments in complex systems under attack by rational and capable adversaries. Such systems include software, hardware, and data critical to national security and industrial competitiveness. Consequently, we believe that QuERIES has wide applicability within both the DoD and private sectors. ■

Acknowledgments

The Office of the Deputy Under Secretary of Defense (Science & Technology) Software Protection Initiative (SPI) sponsored this work. Funding provided via the High Performance Computing Modernization Program. All opinions expressed are those of the authors, not the sponsors.

References

1. W.H. Sanders et al., "Measuring Critical Infrastructure Security," I3P Challenge Description, 2006; www.thei3p.org.
2. S.J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed., Prentice Hall, 2002.
3. J. Wolfers and E. Zitzewitz, "Prediction Markets," *J. Economic Perspectives*, vol. 18, no. 2, 2004, pp. 107-126.
4. D. Geer, K.S. Hoo, and A. Jaquith, "Why the Future Belongs to the Quants," *IEEE Security and Privacy*, vol. 1, no. 4, 2003, pp. 24-32.
5. L. Carin, G. Cybenko, and J. Hughes, "Quantitative Evaluation of Risk for Investment Efficient Strategies in Cybersecurity: The QuERIES Methodology," *Dartmouth Computer Science Technical Report,* 2008.
6. K. Catterjee and W.F. Samuelson, *Game Theory and Business Applications*, Springer, 2001.
7. O. Sheyner et al., "Automated Generation and Analysis of Attack Graphs," *Proc. IEEE Symp. Security and Privacy*, IEEE Press, 2002, pp. 273-284.
8. J. Surowiecki, *The Wisdom of Crowds*, Random House, 2004.
9. R. Hanson, "Combinatorial Information Market Design," *Information System Frontiers 5*, 2003. pp. 107-119.
10. P. Chalasani et al., "A Refined Binomial Lattice for Pricing American Asian Options," *Rev. of Derivatives Research*, vol. 3, no. 1, 1999, pp. 85-105.

***Lawrence Carin** is the William H. Younger Professor of Electrical and Computer Engineering at Duke University. His research interests include signal and image processing, machine learning, and computer security. Carin received a PhD in electrical engineering from the University of Maryland College Park. Contact him at lcarin@ece.duke.edu.*

***George Cybenko** is the Dorothy and Walter Gramm Professor of Engineering at Dartmouth College. His research interests include distributed information, control systems, computer security, and signal processing. Cybenko received a PhD in mathematics from Princeton University. Contact him at gvc@dartmouth.edu.*

***Jeff Hughes** is chief of the ATSPI Technology Office at the Air Force Research Laboratory. His research interests include antitamper systems, cybersecurity metrics, and trust in distributed sensor grids. Hughes received an MS in electrical engineering from the Ohio State University. Contact him at jeff.hughes@wpafb.af.mil.*