Quantitative Evaluation of Risk for Investment Efficient Strategies in Cybersecurity: The QuERIES Methodology

LAWRENCE CARIN
Duke University
GEORGE CYBENKO
Dartmouth College
and
JEFF HUGHES
Air Force Research Laboratory

Information markets, Markov Decision Processes and game theory underlie a new quantitative approach to cybersecurity risk assessment.

1. INTRODUCTION

Organizations in both the private and public sectors have been struggling to determine the appropriate investments to make for protecting their critical intellectual property. As a result, cybersecurity investment strategies at the macro level (overall *strategic* investment in system- or enterprise-wide protection) and the micro level (how to allocate the *tactical* security elements across components of a system or enterprise) have typically been implemented without guidance from a rigorous, quantitative risk assessment and mitigation methodology. Simple questions such as "Are we investing enough?", "What security will have the most impact?" and "How much better is our security now?" are currently difficult to answer [Sanders et al. 2006].

Quantitative Evaluation of Risk for Investment Efficient Strategies (QuERIES) is a novel computational approach to quantitative cybersecurity risk assessment that was designed to answer such questions. It is based on rigorous and quantitative techniques drawn from computer science, game theory, control theory and economics.

Preliminary experiments have corroborated the QuERIES methodology, suggest-

---

Lawrence Carin, Department of Electrical Engineering, Duke University, Durham NC Email: lcarin@ee.duke.edu.; George Cybenko, Thayer School of Engineering, Dartmouth College, Hanover, NH 03755 USA Email: gvc@dartmouth.edu.; Jeff Hughes, AT-SPI Technology Office, Air Force Research Laboratory, Wright-Patterson Air Force Base, OH 45433-7320 Email: jeff.hughes@wpafb.af.mil.

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

ing that it is a broadly applicable alternative to red teaming, black hat analysis, and other decision support methodologies which have previously been tried for cybersecurity related risk assessment.[1]

To date, QuERIES has been focused on the problem of protecting critical Department of Defense intellectual property (IP), in which the loss of one copy of the IP is catastrophic, as opposed to consumer IP, in which the loss of multiple copies can be tolerated as long as sufficient revenue is 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. This focus results in a specific formulation of the attack/protect economic model. In general, however, QuERIES can be applied to other attack/protect scenarios.

To illustrate the QuERIES methodology and how it can be applied in a given software protection context, consider the challenge of assessing the strength of particular 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:

(1) *Model the Security Strategy* - Develop an attack/protect economic model cast in a game theoretic context. Parameters in this economic model represent objective quantities such as the economic value of the IP (the protected software asset) to the IP owner; the cost of developing the IP by an adversary and; the cost of obtaining the IP through other possible means. Other critical parameters of the model relate details of the protection map (a detailed security plan) of the specific protections applied to the IP asset;

(2) *Model the Attacks* - Use the protection map and knowledge of reverse engineering methodologies to build an attack graph represented as a Partially Observable Markov Decision Process (POMDP) [Russell and Norvig 2002] and;

(3) *Quantify Both Models* - Quantify parameters used in both models by performing a controlled red team attack against the protected IP and then using another red or black hat team to conduct an information market [Wolfers and Zitzewitz 2004] for estimating the parameters of the POMDP. Compute the POMDP's optimal policies and feed those into the attack/protect economic model.

Once both models have been evaluated, it is possible to synthesize multiple derived quantities relevant to risk assessment.

For example, given a class of adversaries, 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*, denoted by P_R. This distribution, as explained further in Sections 2.5 and 2.6, assumes that the attacker does not have an a priori model of the attack graph or protection scheme. The attacker is therefore learning the protection scheme through trial and error. The probability distribution is

[1] A "red team" attack involves attackers who have little or no knowledge of a systems' internal protection. A "black hat" analysis involves attackers who have access to design details of the internal protection.

Approved for Public Release: AFRL/WS-07-2145, September 2007

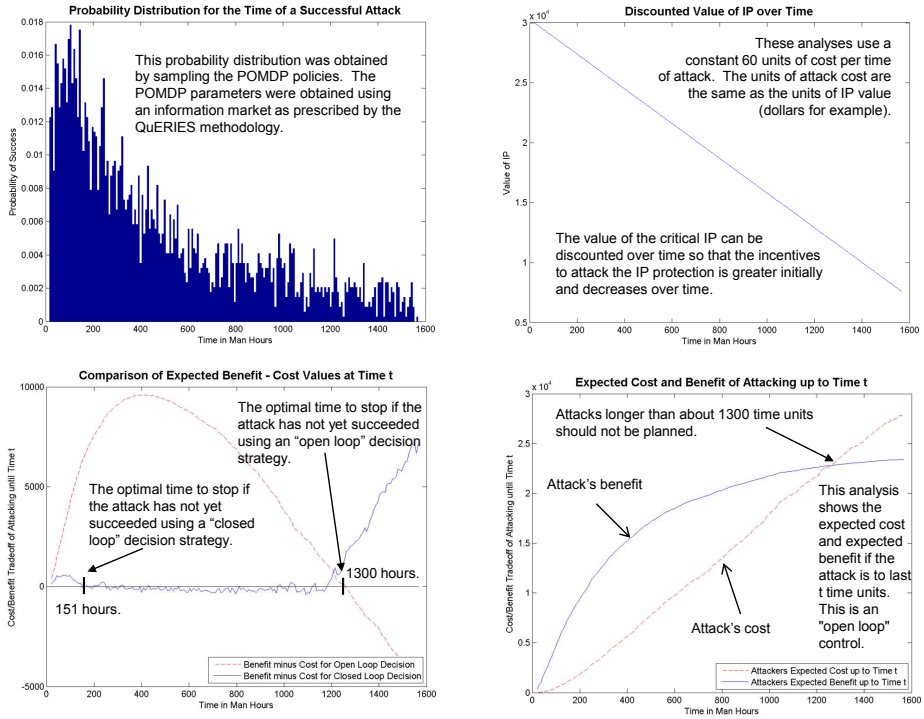


Fig. 1. These plots show various distributions and time-dependent quantities that QuERIES can obtain. These are discussed in more detail in the text.

generated by launching multiple independent attacks under this assumption. In Section 2.6, this same protection scheme will be attacked by an adversary who has insider knowledge and does know the actual optimal attack graph corresponding to the protection scheme.

Figure 1 also depicts some of the possible outcomes and analyses that QuERIES can produce. The top right plot shows the discounted value for the critical IP - its value decreases over time as technology advances and as its mission criticality may diminish. This is an estimate of the value of the IP over time and can be specified by the IP asset owner as appropriate.

The bottom left plot 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 closed loop analysis approach is similar to American Options pricing algorithms used to decide when to exercise an option before its expiration date [Chalasani et al. 1999]. The algorithm for computing that decision strategy is based on dynamic programming and related to classical stochastic control theory. It is a closed loop decision procedure because it uses the fact that an attack has not been successful up to a certain time and factors in the future costs and optimal decisions that will be made in the future. The plot of the closed loop strategy

in the lower left of the figure shows the expected benefit minus cost of continuing the attack given that it has not succeeded up to that time. The other approach depicted in the bottom left of the figure is the open loop analysis which compares the difference between the expected benefit and the expected costs at each time where the policy is to stop the attack when the difference becomes negative.

These two analyses are presented to illustrate two different decision strategies attackers could use. There are a variety of other possible decision strategies. Note however that the results of different analyses could be quite different. Using the closed loop decision algorithm, if the attacker has not succeeded after about 151 hours their optimal decision is to stop the attack because they have 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 $t \leq 151$ plus the expected cost of failure at time $t = 151$) is about \$7895 which can be compared with the \$30,000 initial value of the IP. That is, 0.25*30000 is 7500 which is roughly the expected benefit of the attack up to 151 hours (which is actually 7519) .

The bottom right plot compares 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 continuing to press the attack having passed the "fat" part of the probability distribution and thereby working for a diminishing likelihood of returns. The difference between the benefit and cost in the bottom right plot is shown in the bottom left plot.

The probability distribution P_R that QuERIES obtains can be the basis for different kinds of analyses as this discussion illustrates. Knowledge of P_R before and after certain protections are added or improved can help answer fundamental questions such as: "How much better protected is my IP?"; "What is the right level of investment in its protection?"; "What is the cost/benefit analysis for adding more protections?"

Examples of such before and after analyses are presented in Section 2.6. In addition, it is possible to model attackers who have knowledge of the software protections explicitly as well as attackers who are attacking without such a priori information.

QuERIES therefore is a methodology for addressing two of the key challenges:

—How can the probability distribution, P_R, be effectively and cost-efficiently obtained?

—What are the ways in which P_R can be used for business relevant cybersecurity risk assessment?

Because QuERIES is agnostic about how P_R is actually used most appropriately by a decision maker, 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 [Sanders et al. 2006; Anderson and Schneier 2005; Cybenko 2006]. For example, formal methods (that

is, logical analyses of a design) can only verify that a design has certain desirable properties but are silent on the properties of an actual implementation and 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 will suggest multiple possible attack paths and associated tools required, yet can only offer gross estimates of attack times and costs.

The rest of this paper is organized as follows. Section 2 is a review of the major concepts and steps of the QuERIES methodology. Additional details of related work along with QuERIES results, derivations and experiments are documented in reports available from the authors.

2. THE QUERIES METHODOLOGY

The steps in the QuERIES methodology are described next. As previously noted, the three major elements of the methodology are:

(1) *Model the Security Strategy*;
(2) *Model the Attacks* and;
(3) *Quantify Both Models.*

These components are broken down into a series of 7 steps that are illustrated in Figure 2.

2.1 Identify Critical IP Assets and Threats Against Them

Users of the QuERIES methodology must first identify their critical IP assets and the threats against them through analysis of their various missions or strategic plans. In general, critical IP assets can be found in hardware, software or data and have the following characteristics:

(1) the IP embodies knowledge and information obtained at significant cost to the IP owner;
(2) an adversary desires to obtain the IP;
(3) an adversary would value the IP at roughly the same level as the asset owner does.

The fact that IP was expensive for the owner to obtain or develop is not in and of itself enough to characterize it as critical. QuERIES in particular assumes that IP under consideration is desired by the adversary. We also assume that the IP has already been developed and so the cost of developing it is sunk and cannot be recovered by the owner.

We will be using a relatively objective measure of the value of such an asset, namely the cost to develop it. Those costs can usually be estimated relatively reliably using programmatic information although in many cases the development of advanced systems leverages a broad technology base that may have been already expensed elsewhere. Our notation for the owner's cost of developing the IP is denoted by C_{IP}.

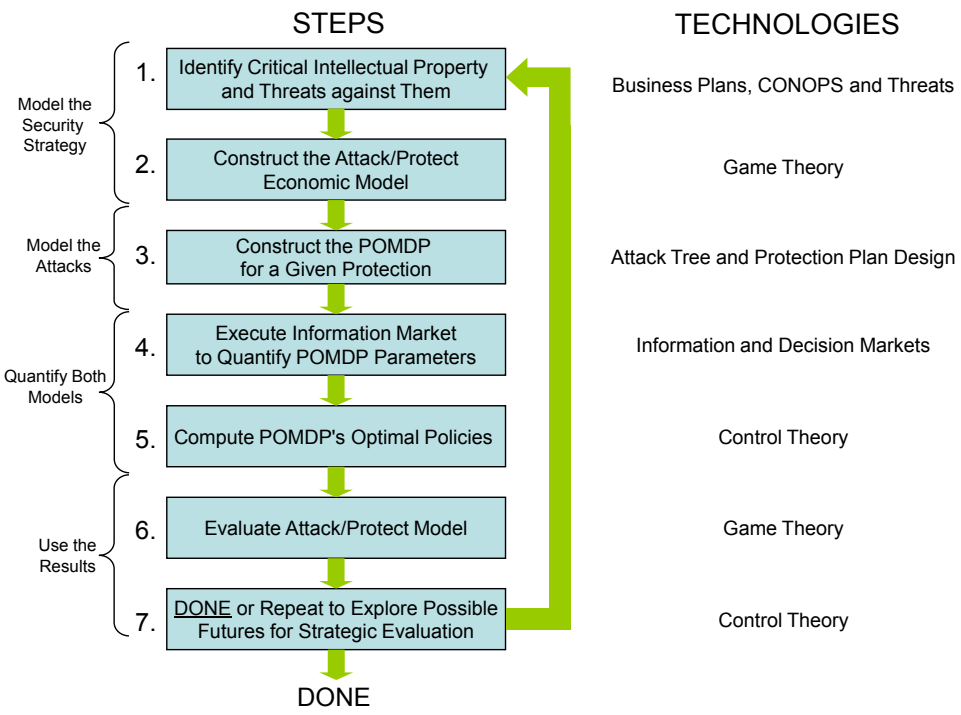


Fig. 2. The steps of the QuERIES methodology and their supporting technologies. These technologies are established and well understood. QuERIES combines them in a novel way.

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. Advances in software development technology, basic science, semiconductor fabrication processes and other factors could make development today less expensive for example.

Threats against critical IP could include careless handling of the IP by insiders with access, deliberate mishandling of the IP by malicious insiders, stealing unprotected copies of the IP, stealing protected copies of the IP and reverse engineering stolen copies of the IP.

Hence the first step of the QuERIES method identifies:

—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 ab initio;

—P_S : the probability of stealing the unprotected IP based on historical data for

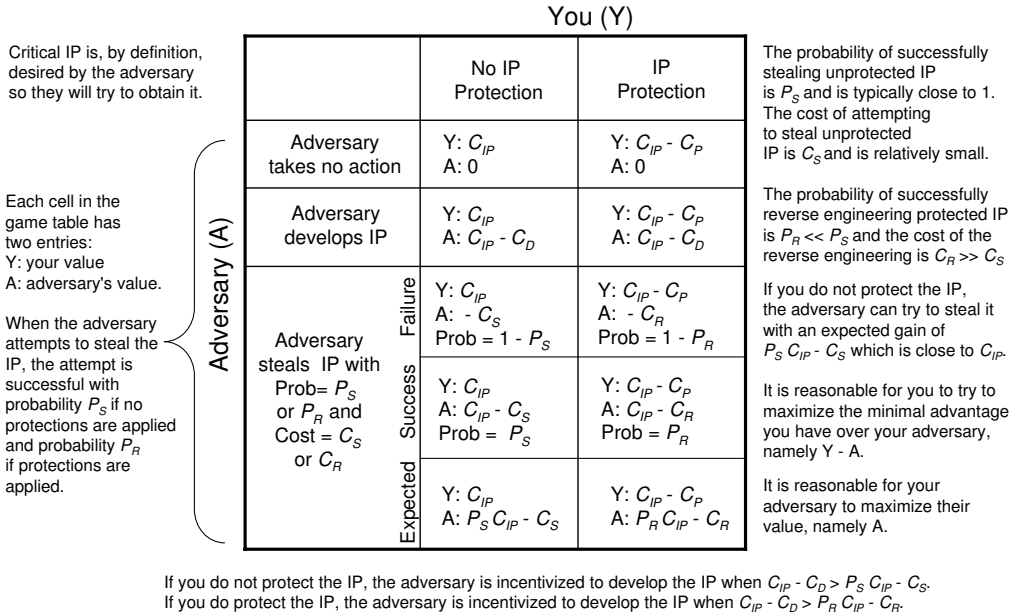


Fig. 3. In this example, the QuERIES economic model is based on a simple game theoretic formulation. In the game, the IP owner can protect or not protect and the adversary can develop the IP ab initio or attempt to steal or reverse engineer the IP. Although the case in which the adversary chooses to do nothing is listed, the definition of critical IP is that they will try to obtain the IP. The consequences of various moves are discussed in both the figure commentary and text of the paper.

similar IP for example;

—C_S : the cost of stealing the unprotected IP based on historical data for similar IP.

These quantities could be estimated for different adversaries who have different technology bases from which to recreate the IP and different capabilities for stealing the unprotected IP.

2.2 Construct the Attack/Protect Economic Model

The QuERIES attack/protect economic model is based on a game with two players - the protector (you) and the attacker (your adversary). Game theory is a mature discipline which was originally rigorously developed to support strategic, military and policy decision making [von Neumann and Morgenstern 1944; Isaacs 1999]. It has subsequently been extended and widely used for business and economic applications as well [Catterjee and Samuelson 2001].

The two basic game moves available to the protector (you) are: protect critical IP or not protect IP. The game tables shown in Figure 3 shows the protector's moves in the columns labeled "No IP Protection" and "IP Protection." Different protection technologies are possible for a given IP so that in practice several moves

are possible for the protector, one for each type of protection being considered.

In this example, we are modeling 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. "No Action" is not a viable move. For each combination of moves by the protector and attacker, we write down an expression for the resulting loss or gain in the corresponding cell of the game table. When an adversary attempts to steal or reverse engineer critical IP, there is some probability that they will succeed, namely P_S and P_R respectively.

The QuERIES analysis of the game is based on the following player objectives. The IP asset owner wants to maximize the minimal advantage they have 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 their value without factoring in the owner's value because they want the IP and the owner already has it. The IP asset owner moves first because they get to decide whether to deploy protections or not when the critical IP is fielded.

If the IP owner (you) does not protect the IP, the adversary will attempt to steal the IP for realistic parameter values. Specifically, we assume the probability of successfully stealing the IP, P_S is close to 1 and 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 is typically about the same as C_{IP}. That is, the cost to the adversary of developing the IP ab initio is about the same as the cost to the owner of developing the IP originally.

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 ab initio otherwise.

The IP owner must take into account their cost of protecting the IP so as not to have a Pyrrhic victory by making the cost of protection more expensive to them than the cost of obtaining the IP for their adversary.

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

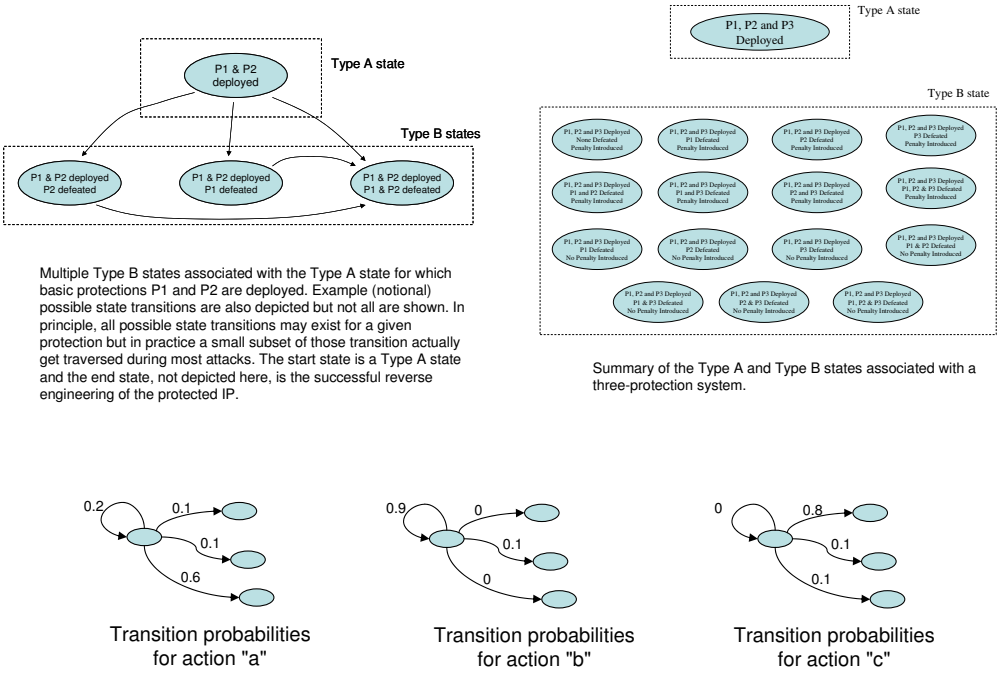
$$C_D - C_P > (1 - P_S) * C_{IP} + C_S$$

and

$$(1 - P_R) * C_{IP} + C_R - C_P > (1 - P_S) * C_{IP} + C_S.$$

The above 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 adversary's expected value if they choose to steal the IP, $P_S * C_{IP} - C_S$, which we have seen they will do in realistic situations. The left sides of the inequalities are the IP owner's relative advantages if the owner chooses to protect for the cases that the adversary develops or reverse engineers the IP, respectively.

This economic analysis requires several quantities: C_P, C_D, P_S, C_S and C_{IP} which can be estimated from available empirical data. The quantities which cannot be readily estimated from historical data are P_R and C_R. Estimating these quantities and related derivatives is the objective of the next few steps of the QuERIES



Multiple Type B states associated with the Type A state for which basic protections P1 and P2 are deployed. Example (notional) possible state transitions are also depicted but not all are shown. In principle, all possible state transitions may exist for a given protection but in practice a small subset of those transition actually get traversed during most attacks. The start state is a Type A state and the end state, not depicted here, is the successful reverse engineering of the protected IP.

Summary of the Type A and Type B states associated with a three-protection system.

Fig. 4. A POMDP formulation of attacks against protected IP requires specification of the POMDP structure. Two example structures are depicted in the top of this figure. In addition to the underlying structure, the specification of a POMDP requires transition probabilities corresponding to the various action possible, costs and observables. The bottom part of the figure depicts examples of different transition probabilities that correspond to three different actions taken in a state.

methodology.

2.3 Construct the POMDP

The purpose of the present and the next two steps is to obtain 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. As discussed previously, these quantities are fundamental to risk assessment questions but have been difficult to obtain previously.

The QuERIES' approach to estimating the probabilities and costs of successful attacks begins by first formulating the possible attacks based on the given protection map in terms of a graph that represents the adversary's multi-stage decision processes and states of knowledge. Attack graphs have been used effectively in computer security studies [Sheyner et al. 2002] and QuERIES advances the concept in several fundamental ways. Perhaps the most fundamental contribution that QuERIES makes to cybersecurity attack modeling is the introduction of partially observable Markov decision processes (POMDP) [Kaelbling et al. 1998] to the field. A POMDP is a powerful tool for modeling a single agent operating in an en-

vironment. POMDP's have been used successfully in a variety of applications, the most notable recent example being the 2005 DARPA Autonomous Vehicle Grand Challenge [Thrun 2006]

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 towards 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 may or may not require defeating all of the protections used.

An attacker takes actions while attempting to defeat the protections that have been 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, the probability of transiting from one state to another is modeled by a Markov process, specifically a finite Markov chain in this case. That is, for every action, there is a Markov chain labeled by that action that specifies the state transition probabilities resulting from that action. Another consequence, that depends on the action taken and the current state, is that a cost is incurred.

Such a modeling formalism is called a Markov Decision Process (MDP). MDP's have been comprehensively studied in the context of control theory, operations research and economics for many years. This has resulted in many theoretical results and computational algorithms for MDPs [Feinberg and Shwartz 2002].

While MDP's provide the basic modeling formalism, they are not quite enough. The problem is that at any given time, an attacker typically does not know what state he is in (the states are not directly "observable"). In the QuERIES context, the attacker may not know what protections have been deployed, which protections have been defeated, and is uncertain about what penalties may have been introduced through his previous actions. The attacker does have access to certain observations, which are a stochastic function of the last action taken and the current state occupied.

MDP's in which states are not directly observable can be modeled as *Partially Observable Markov Decision Processes* or POMDPs [Kaelbling et al. 1998]. The reader is referred to the extensive literature on MDP's and POMDP's for details of those technologies [Feinberg and Shwartz 2002; Kaelbling et al. 1998; Thrun 2006].

The output of this step of the QuERIES methodology is a POMDP structure which 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 complete specification of the POMDP 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 which is not directly observable by the attacker. These quantities are estimated through controlled red team attacks combined with subsequent information or decision markets described in the next step.

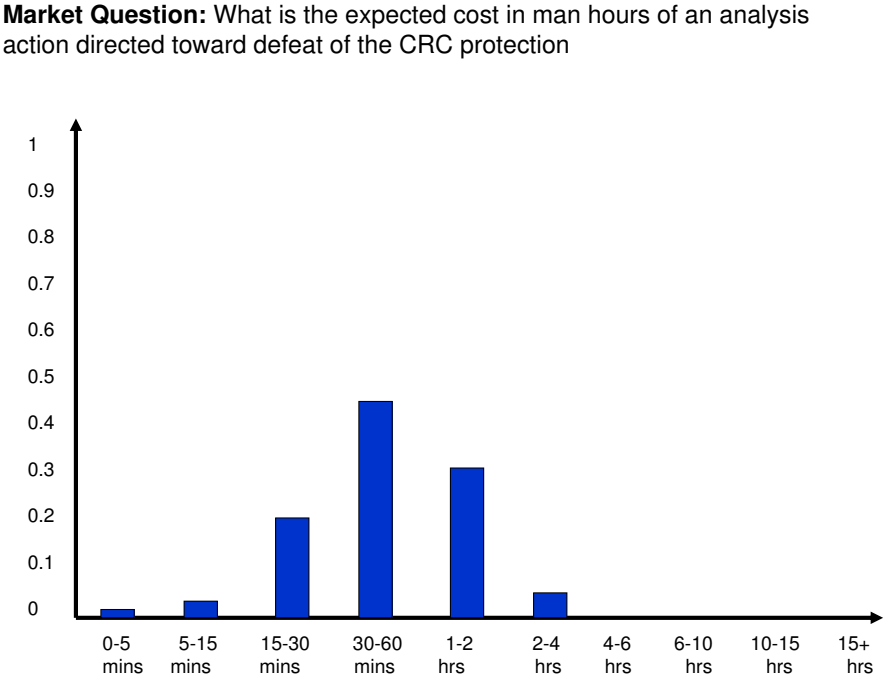


Fig. 5. This figure shows the results of using an information market with market scoring rules to estimate a probability distribution collectively by a second red team or black hat group. Market scoring is a recently developed mechanism that is effective for estimating distributions over large configuration spaces.

2.4 Execute Information Markets

A fundamental challenge in cybersecurity risk assessment has always been estimating the probability and cost of successful attacks against proposed and deployed security technologies. Traditional approaches to risk assessment have been to conduct red team and black hat exercises. This type of actual or virtual exploitation does not provide enough information to obtain quantitative probability distributions or cost estimates.

The QuERIES economic model for IP protection previously described requires these probabilities and costs as well. However, the approach QuERIES uses is quite different and we believe far more powerful than red team or black hat attacks alone.

As described in the previous step, QuERIES represents attacks using partially observable Markov decision processes (POMDP). At first blush, 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. A fundamental insight of QuERIES is that the underlying POMDP parameters can be obtained from red teams but not in the traditional way. Using information or decision markets, it is possible to estimate with high

accuracy the underlying POMDP parameters and then, as described in the next step, compute optimal policies for the POMDP and then perform simulations on those policies to obtain the probability distribution of P_R.

Information markets are mechanisms designed for participants to interact with each other using simple exchanges of structured information. The outcome of an information market is a collective, not consensus, estimate of a quantity. Information markets have recently received a great deal of attention in the popular press [Surowiecki 2004] and technical literature [Hanson 2003; Gneiting and Raftery ; Roll 1984; Chen and Plott]. They are increasingly being used in business to forecast sales, market trends and complex system behaviors.

Examples of information markets are traditional financial markets like stock and commodities exchanges in which participants buy and sell shares, options and various derivatives. The only information effectively exchanged by participants in those markets are 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 value of the instrument, such as the valuation of a company or the future prices of commodities such as oil, orange juice or lumber products [Roll 1984].

Pari-mutuel betting, such as at horse race tracks, also involves the exchange of information only through the tote board odds for a race. The odds have a natural, intrinsic interpretation as probabilities for how the different horses will perform.

Information markets have been also been established for political races, current events, financial news, weather and unique events [IEM ; Intrade.com]. Their effectiveness, if properly constituted, is not controversial today [Surowiecki 2004]. The proper design of mechanisms for estimating probabilities, probability distributions and other quantities of interest is a small but rapidly growing industry [HSXResearch]. Mechanisms have been discovered to effectively estimate probability distributions over large combinatorial spaces in which the number of individual elements can be extremely large [Hanson 2003; Gneiting and Raftery].

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

Another red team, different from the market red team ideally, then actually conducts a traditional red team exercise to determine what is "correct" in order that payouts can be determined. The fundamental outcome of the market are the estimates of probability distributions and costs, not the specific traditional red team exercise outcome. Using horse racing as an analog, running a race once merely generates a single sample from the probability distribution over the horses about which one will win. The information market premise is 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.

Approved for Public Release: AFRL/WS-07-2145, September 2007

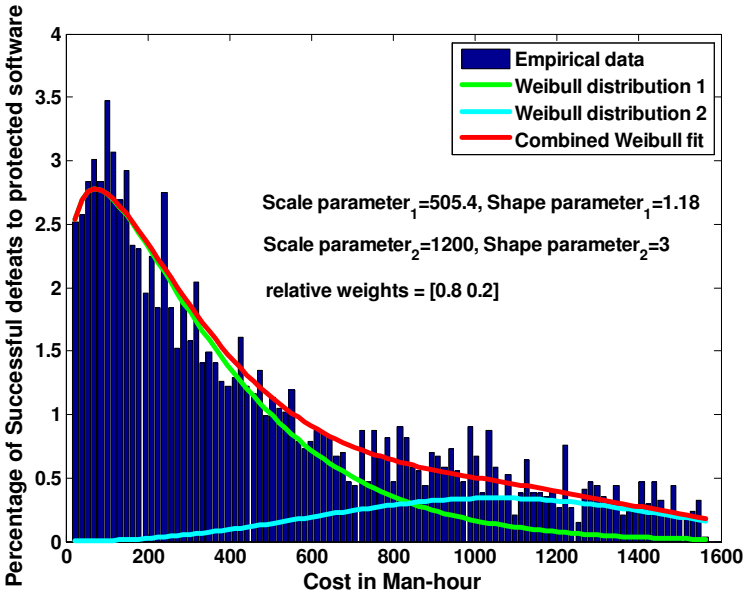


Fig. 6. The graph show a probability distribution for the time to break critical IP protections through reverse engineering obtained from the QuERIES methodology. A mixture of Weibull distributions has been fit to the data for illustrative purposes.

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

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. In fact, the various quantitative results shown in this paper were obtained by the QuERIES methodology using actual red team information markets.

2.5 Compute POMDP's Optimal Policies

Once the QuERIES information market has produced estimates of the POMDP's probabilities and costs, standard techniques for finding optimal policies of POMDP's can be used to determine the optimal action to take in each state to minimize a cost objective [Kaelbling et al. 1998]. A common objective to optimize is cost which, in the case of protecting critical IP, can be measured in time to defeat the protections.

The optimal policy prescribes the action to take in each state of the POMDP to minimize the expected cost/time. A simulation of the POMDP that produces multiple runs using the optimal policy in every step will generate an empirical prob-

ability distribution of times to defeat the protections assuming that the attacker has knowledge of the optimal attack policy. Figure 1 showed, by contrast, 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 optimal policy of the POMDP before starting the attack.

By sampling from the policy space, it is possible to generate empirical distributions corresponding to suboptimal policies which represent less skilled or capable adversaries. For example, it is possible to randomly sample the second and/or third best policies in addition to the optimal policy 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.

2.6 Evaluate Attack/Protect Model

The previous steps of the QuERIES methodology have produced optimal and suboptimal attack policies which can be used to generate a variety of probability distributions for the time (and therefore cost) of successfully defeating the protections applied to critical IP.

The probability distribution P_R can be the basis for different kinds of analyses. Possible risk-related derivatives of P_R are, for example:

—The expected cost of defeating the protection:

$$\sum_{i=0}^{\infty} c_i P_R(i)$$

where c_i is the cost of the ith man-hour in the attack;

—The expected time to defeat the protection:

$$\sum_{i=0}^{\infty} i P_R(i);$$

—The expected cost of defeating the protection given that the protection is defeated at or before time t:

$$\frac{\sum_{i=0}^{t} c_i P_R(i)}{\sum_{i=0}^{t} P_R(i)}$$

—The optimal decision time for an attacker to quit if they have not yet succeeded;

—The associated probabilities of success, costs and time under the above optimal decision policy.

As illustrated previously in Figure 1, there are "open loop" and "closed loop" strategies for executing attacks assuming that an attacker has some knowledge of P_R. The open loop strategy does not take into account the fact that an attack has not succeeded as it progresses. The closed loop strategy does and the algorithmic basis for computing this strategy is similar to pricing an American-style financial option [Chalasani et al. 1999]. If an attacker does not know P_R, the attacker's strategy will be based on some historical experience that they have with cyber

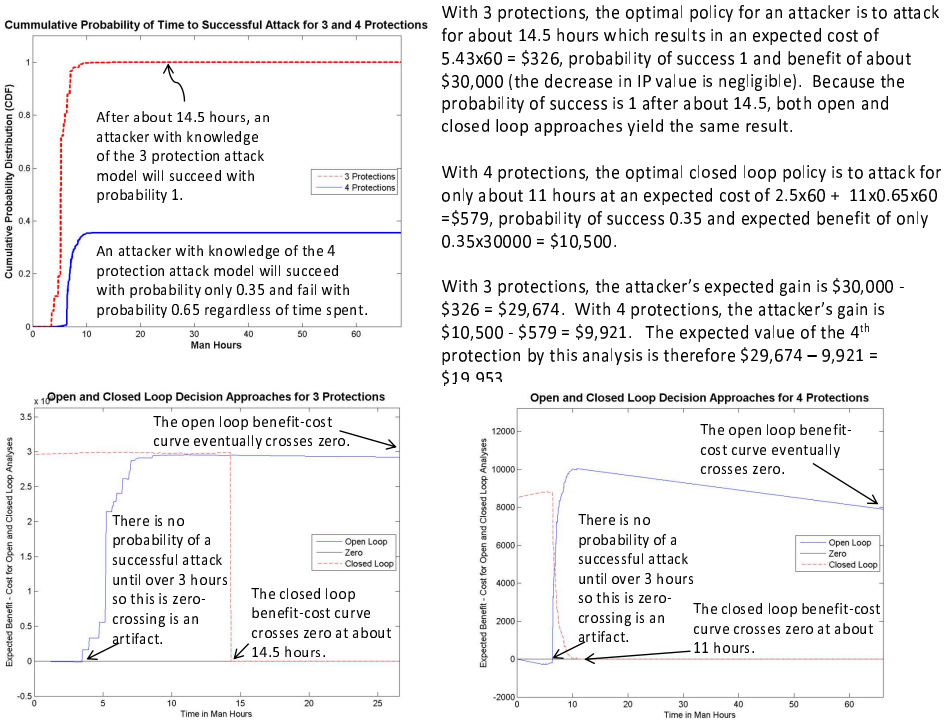


Fig. 7. This figure shows an analysis that can be made to compare two protections and the values of the protections relative to each other.

protections. We believe that different attack policies and associated derivative risk assessment quantities should be explored more extensively in the cybersecurity and critical IP protection domains.

Knowledge of P_R before and after certain protections are added or improved can help answer fundamental questions such as: "How much better protected is my IP?"; "What is the right level of investment in its protection?"; "What is the cost/benefit analysis for adding more protections?"

To illustrate the possible comparisons between protections that can be made, we have performed a worst-case analysis (for the protector of the IP) based on the data shown in Figure 1 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 expends much effort exploring the protection and attack space (that is, the attacker does not employ the optimal attack policy). The resulting probability distribution of time to defeat was previously shown in Figure 1.

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 then two orders of magnitude smaller as the plot in the top left corner of Figure

7 shows. We can repeat this analysis if another layer of protection is added, also as shown in the top left corner of Figure 7. With 4 protections, successful attacks were only found with probability 0.35 so that with probability 0.65 an attack will be unsuccessful in any reasonable time.

The bottom two plots in Figure 7 compare the "open loop" and "closed loop" benefit minus cost curves and associated stopping times for 3 protections in the bottom left and for the 4 protection case on the right. Subsequent analysis is discussed on the top right corner of the figure. The point is that even though successful attacks are possible 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 3 protections, the optimal policy for an attacker is to attack 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 5.43x60 = $326 for a successful attack and a benefit of about $30,000 (the decrease in IP value is negligible).

With 4 protections, the optimal policy is to attack for only about 11 hours with a probability of success of only 0.35. The expected number of hours prior to the 11 hour stopping time is 2.5. With probability 0.65, the attacker will expend those 11 hours but fail. Therefore the expected cost is 2.5x60 + 11x0.65x60 =$579, and expected benefit is only 0.35x30000 = $10,500.

It should be noted that the reason there is only a 0.35 probability of successfully attacking the 4 protections is that we are using a bounded, fixed resource model for the attacker. That is, the representational capacity and computing power of the adversary are the same for 3 and 4 protections. With 4 protections, those resources are not sufficient to conduct successful attacks more than 35% of the time an attack is attempted. This is a resource limitation that arises intrinsically when numerically computing optimal policies for POMDP's, the details of which are beyond the scope of this paper. (Briefly, the same sized approximation is used to represent the probability distributions underlying both 3 and 4 protection POMDP models.)

In any case, as a result, we can conclude that with 3 protections, the attacker's expected gain is $30,000 - $326 = $29,674 while with 4 protections, the attacker's gain is $10,500 - $579 = $9,921. Therefore, the added "insurance" of the 4th protection is to safe guard $29,674 - 9,921 = $19,953 of the IP value.

3. RELATED WORK

3.1 Cybersecurity Risk Assessment

The general subject of risk assessment is an established discipline with a long history in both the military and commercial sectors [Bernstein 1998; Boehm 1991]. It is especially mature in the financial and insurance industries where economic models of risk can be based on accepted theory and large amounts of historical data are available. However, the absence of a theoretical framework and actuarial-class data about information assurance makes risk assessment, mitigation and management a major challenge in the computer security domain today [Daniel Geer et al. 2003; Cybenko 2006].

For example, previous metrics for cybersecurity risk assessment, such as they are

presently known, have proven to be inadequate for a variety of reasons including [Anderson and Schneier 2005; Sanders et al. 2006]:

(1) Most technical metrics are not quantitative: in many cases, metrics are associated with how closely organizations follow specified processes;
(2) Most metrics are lagging as opposed to leading indicators of performance: existing metrics are not useful for predictive uses;
(3) Metrics with different objectives are not integrated to provide a comprehensive view: organizational, technical and operational security metrics have not been successfully combined to provide overall security assessments;
(4) Metrics are not absolute: they measure quantities that may not be relevant to specific missions and institutional goals;
(5) Metrics based on formal methods are powerful but not sufficient: they are based on assumptions which are typically difficult or impossible to verify in real operational settings.

By constrast, the QuERIES methodology produces business-relevant quantitative metrics that address the above concerns.

4. SUMMARY

Consequences of the QuERIES approach include:

(1) Improved threat characterization including methodologies to obtain statistical parameterizations of the *global red team*;
(2) A methodology for evaluating IP protection schemes during the design phase, as well as when first fielded, with an ability to identify the weakest links, and to perform a cost-benefit analysis for strengthening the protections most appropriately;
(3) A predictive methodology for the evaluation of protection schemes *over time* which allows tracking the evolution of the dynamic protector-attacker game theory model and quantifies the impact of the attacker's learning curve on protection effectiveness;
(4) The ability to link these quantitative risk assessments to an organization's strategic objectives and business plan via an economic model.

In other words, the QuERIES methodology can be used 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 significant and immediate impact on the protection of critical IP, including weapons systems designs, chip designs, complex computer software and databases containing personal and financial information.

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

Approved for Public Release: AFRL/WS-07-2145, September 2007

REFERENCES

Anderson, R. and Schneier, B. January 2005. Guest editors' introduction: Economics of information security. *IEEE Security and Privacy 3,* 1, 12 – 13.

Bernstein, P. L. 1998. *Against the Gods: The Remarkable Story of Risk.* Wiley and Sons, New York.

Boehm, B. Jan. 1991. Software risk management: principles and practices. *IEEE Software 8,* 1, 32 – 41.

Catterjee, K. and Samuelson, W. F. 2001. *Game Theory and Business Applications (International Series in Operations Research & Management Science).* Springer, New York, NY.

Chalasani, P., Jha, S., Egriboyun, F., and Varikooty, A. January, 1999. A refined binomial lattice for pricing American Asian options. *Review of Derivatives Research 3,* 1, 85–105.

Chen, K.-Y. and Plott, C. Information aggregation mechanism: Concept, design and implementation for a sales forecasting problem. California Institute of Technology, Division of the Humanities and Social Sciences Working Paper, Number 1131.

Cybenko, G. 2006. Why Johnny cańt evaluate security risk. *IEEE Security and Privacy 4,* 1, 5.

Daniel Geer, J., Hoo, K. S., and Jaquith, A. July 2003. Why the future belongs to the quants. *IEEE Security and Privacy 1,* 4, 24 – 32.

Feinberg, E. A. and Shwartz, A. 2002. *Handbook of Markov Decision Processes.* Kluwer, New York.

Gneiting, T. and Raftery, A. Strictly proper scoring rules, prediction, and estimation. Tech. Report 463, Dept. of Statistics, Univ. Washington, Sept. 2004.

Hanson, R. 2003. Combinatorial information market design. *Inf. System Frontiers 5*, 107–119.

HSXResearch. Virtual markets. http://www.hsxresearch.com/.

IEM. The Iowa Electronic Markets. http://www.biz.uiowa.edu/iem/.

Intrade.com. http://www.intrade.com/.

Isaacs, R. 1999. *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization.* Courier Dover Publications, New York, NY.

Kaelbling, L., Littman, M., and Cassandra, A. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence 101*, 99–134.

Roll, R. Dec. 1984. Orange juice and weather. *Am. Economic Rev. 74*, 861–880.

Russell, S. J. and Norvig, P. 2002. *Artificial Intelligence: A Modern Approach (2nd Edition).* Prentice Hall, New York, NY.

Sanders, W. H. et al. March 7, 2006. Measuring critical infrastructure security. I3P Challenge Description, www.thei3p.org.

Sheyner, O., Haines, J., Jha, S., Lippmann, R., and Wing, J. M. 2002. Automated generation and analysis of attack graphs. In *Proceedings of the IEEE Symposium on Security and Privacy.* 273– 284.

Surowiecki, J. 2004. *The Wisdom of Crowds.* Random House.

Thrun, S. 2006. Winning the DARPA grand challenge. In *Proceedings of KDD 2006.* 4.

von Neumann, J. and Morgenstern, O. 1944. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, NJ.

Wolfers, J. and Zitzewitz, E. Spring 2004. Prediction markets. *Journal of Economic Perspectives 18,* 2, 107–126.

Approved for Public Release: AFRL/WS-07-2145, September 2007