**Final Technical Report**

# A Game Theoretic Approach to Cyber Attack Prediction

Award No.: DE-FG02-02ER25527
Duration: 08/16/2003 – 08/15/2005

By

Peng Liu
School of Information Sciences and Technology
Pennsylvania State University
313G IST Building
University Park, PA 16802

Prepared for:

Mathematical, Information, and Computational Sciences Division
Office of Advanced Scientific Computing Research
Office of Science
Department of Energy
Germantown, MD 20874-1290

November, 2005

## Disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of author(s) and do not necessarily reflect the views of the Department of Energy.

# A. Executive Summary

The area investigated by this project is cyber attack prediction. With a focus on correlation-based prediction, current attack prediction methodologies overlook the *strategic nature* of cyber attack-defense scenarios. As a result, current cyber attack prediction methodologies are very limited in predicting strategic behaviors of attackers in enforcing nontrivial cyber attacks such as DDoS attacks, and may result in low accuracy in correlation-based predictions.

This project develops a game theoretic framework for cyber attack prediction, where an automatic game-theory-based attack prediction method is proposed. Being able to quantitatively predict the likelihood of (sequences of) attack actions, our attack prediction methodology can predict fine-grained strategic behaviors of attackers and may greatly improve the accuracy of correlation-based prediction. To our best knowledge, this project develops the first comprehensive framework for incentive-based modeling and inference of attack intent, objectives, and strategies; and this project develops the first method that can predict fine-grained strategic behaviors of attackers.

The significance of this research and the benefit to the public can be demonstrated to certain extent by (a) the severe threat of cyber attacks to the critical infrastructures of the nation, including many infrastructures overseen by the Department of Energy, (b) the importance of cyber security to critical infrastructure protection, and (c) the importance of cyber attack prediction to achieving cyber security.

## B. Comparison with the Goals and Objectives of the Project

Compared with the goals and objectives of the project:

- In general, we have accomplished the primary goal of this project, that is, to develop a game theoretic methodology for cyber attack prediction.

- In particular, we have developed a comprehensive game theoretic framework for modeling and inference of attack intent, objectives and strategies (AIOS), which builds the foundation for the attack prediction methods we proposed in both **Task I** and **Task II**, since accurate attack predictions against *intelligent* and *active* attackers are very difficult, if not impossible, to produce if the attacker's intent, objectives and strategies are neglected. Furthermore, several important research issues we proposed in the original proposal, such as (a) "using signaling games to predict attacks" (in Section 1.3.7), (b) "subgame-perfect Nash equilibrium strategies" (in Section 1.3.8), and (c) "predicting simultaneous attacks from multiple attacks" (in Section 1.3.9), are all addressed in a systematic way within this framework.

- As a primary objective of **Task I.A: Predicting attacks that can only be detected by anomaly detection**, we have developed a game theoretic approach to predict, detect and analyze financial cyber crimes such as credit card frauds and money laundering.

- As a primary objective of **Task I.B: Predicting DDoS attacks**, we developed a game theoretic approach to predict DDoS attacks against a network armed with the Pushback defense (developed by AT&T), and we have done extensive ns2 simulation experiments (above 5000 experiments) to produce concrete DDoS attack predictions and justify the benefits of this approach.

- We have adjusted the original objective of **Task I.C: Predicting other known types of cyber attacks** from predicting attacks on firewall-protected systems and predicting buffer overflow attacks to applying the game theoretic AIOS modeling and inference framework (developed by us) to measure the resilience of computer systems (and networks) against cyber attacks. In particular, we have built a game theoretic framework to measure the resilience of generic computer systems against cyber attacks, and we have done extensive ns2 simulation experiments to measure the Internet's resilience against DDoS attacks.
    - We have done the adjustment partially because we found during the research that measuring the resilience of computer systems against cyber attacks is a natural "application" of the AIOS modeling and inference framework, and from a research viewpoint applying the developed AIOS modeling and inference framework to measure the resilience of computer systems against cyber attacks is a more challenging, more interesting, more important problem than the original objective of Task I.C. Note that after the comprehensive AIOS modeling and inference framework is developed, predicting attacks on

firewall-protected systems and predicting buffer overflow attacks become a quite straightforward application of the framework.

- We have adjusted our objectives of **Task II: Predicting new types of attacks** to first investigate (predictive) defenses against data (and code) corruption attacks in such information systems as workflow systems, then investigate how to build game theoretic prediction models against these unknown or new attacks.
    - We have done the adjustment due to the following reason: To successfully perform **Task II.A**, namely predicting unknown or new types of data corruption attacks, we need to first investigate (predictive) defenses against data (and code) corruption attacks in such information systems as workflow systems before investigating how to build game theoretic prediction models against these unknown or new attacks. The rational is that good attack prediction models cannot be developed without a good understanding of both the attacks and the corresponding defenses.

- Partially funded by the grant, we have extended our research scope to investigate secure information sharing in collaborative information processing, intrusion masking distributed computing, and DDoS attacks in mobile ad hoc networks.

# C. Project Activities

As indicated in Section B, the project activities can be summarized as follows in terms of five research themes: (1) Modeling and Inference of Attacker Intent, Objectives and Strategies (AIOS); (2) Prediction and Analysis of Internet DDoS Attacks; (3) Prediction, Detection and Analysis of Financial Cyber Crimes; (4) Predicting the Resilience of Computer Systems against Attacks; (5) Attack Resistant Workflow Systems.

## C.1 Modeling and Inference of Attacker Intent, Objectives and Strategies

As a fundamental component of **Task I and Task II**, we developed a game theoretic framework for the modeling and inference of Attacker Intent, Objectives and Strategies (AIOS). The results are summarized in the following papers:

- P. Liu, W. Zang, M. Yu, "Incentive-Based Modeling and Inference of Attacker Intent, Objectives and Strategies", *ACM Transactions on Information and Systems Security,* Vol. 8, No. 1, 2005, 78-118.
- P. Liu, W. Zang, "Incentive-Based Modeling and Inference of Attacker Intent, Objectives and Strategies," *Proc. 10th ACM Conference on Computer and Communications Security (CCS '03)*, October 28-31, Washington DC, 2003, pages 179-189.

[**Motivation**] Predicting cyber attacks is very critical for cyber homeland security. One fundamental weakness of existing cyber security technologies is that they can only *passively* prevent, detect, and react to cyber attacks. As a result, no cyber system can prevent all attacks; intrusion detection always lags behind the attacks; and intrusion response in many cases is "too late" after very serious damage is caused. The ability to predict cyber attacks in a timely fashion with precision can transform passive to proactive cyber defense, where the attack will be "contained", "isolated", or "defeated" before it causes deadly damage.

Unfortunately, existing cyber attack prediction techniques, which are *attack-oriented*, are very limited. Attack-oriented prediction uses the attacks that have already happened to predict new attacks. In a dynamic attack-defense scenario, alerts of intrusions are correlated based on the *causal relationships* between them to build attacking scenarios and infer subsequent attacking actions. However, attack-oriented prediction has two serious drawbacks: (1) before a multi-step cyber attack is launched, nothing about the attack can be predicted. (2) During an earlier stage of a multi-step attack, the *postconditions* of one (detected) action of the attack may match the *preconditions* of a large number of subsequent attacking actions; and as a result, alert correlation can do nothing except waiting until a more complete attack scenario emerges (since alert correlation cannot tell which subsequent action is more likely to be taken by the *attacker*). However, such waiting may lead to "too late" threat response.

[**Hypothesis**] Attack-oriented prediction is so limited in proactive cyber defense primarily because they overlook attacker *motives* or *intents*. With the ability to infer attacker motives, even at an early stage of an attack, we may still be able to *infer* or *predict* what the attacker is intending to do.

[**Approach**] Accordingly, we developed an *attacker-oriented* cyber attack prediction framework, where attacker motives are modeled and exploited to infer attacks, and the drawbacks of attack-oriented prediction can be overcome. The framework consists of an incentive-based method to model AIOS and a game theoretic approach to infer attack *strategies*. On one end, we found that the concept of (economic) *incentives* can unify a large variety of attacker intents; the concept of *utilities* can integrate incentives and *costs* in such a way that attacker objectives can be practically modeled. On the other hand, we developed a game theoretic AIOS formalization (as visualized in Figure 1) which can capture the inherent *inter-dependency* between AIOS and defender objectives and strategies in such a way that attack strategies can be automatically inferred based on the *equilibrium* strategies of the *game* between the cyber system and the attacker.

Accordingly, we formalized two types of AIOS inference problems based on the AIOS formalization, which are illustrated below:

**Type A - Infer Attack Strategies**. Given a specific model of attack intent and objectives, infer which attack strategies are more likely to be taken by the attacker. Our framework implies the following pipeline in inferring attack strategies:

(1) Make assumptions about the system and the (types of) attacks that concern the system. Note that practical attack strategies inferences may only be able to be computed within some domain or scope (due to the complexity).
(2) Model the attacker intent, objectives and strategies (conceptually). Specify the attacker's utility function and strategy space. Estimate the attacker's knowledge base.
(3) Specify the system's metric vector and security vector. Specify the system's utility function and strategy space. Build the system's knowledge base.
(4) Determine the game type of the game theoretic attack strategy inference model that will be developed, then develop the model accordingly.
(5) Compute the set of Nash equilibrium strategies of the attack strategy inference game model developed in Step 4. A key task is to handle the computation complexity. If the *complexity* is too much, we need to do (inference) precision performance tradeoffs properly using some (semantics-based) approximate algorithms.
(6) Validate the inferences generated in Step 5. The relevant tasks include but are not limited to *accuracy analysis* (i.e., how accurate the inferences are) and *sensitivity analysis* (i.e., how sensitive the inferences are to some specific model parameters). The relevant validation techniques include but are not limited to (a) investigating the degree to which the inferences match the real world intrusions; (b) extracting a set of high-level properties or features from the set of inferences and asking security experts to evaluate if the set of properties match their experiences, beliefs, or intuitions.
(7) If the validation results are not satisfactory, go back to Step 1 to rebuild or improve the inference model.

**Type B - Infer Attacker Intent and Objectives**. Based on the attack actions observed, infer the intent and objectives of the attacker in enforcing the corresponding attack. To a large degree, the pipeline for inferring attacker intent and objectives is the reverse of that for

inferring attack strategies. In particular, the pipeline has two phases: the *learning* phase and the *detection* phase, which are as follows.

- In the learning phase, do the same thing in Step 1 as the previous pipeline. In Step 2, identify and classify the possible models of attacker intent and objectives into a set of representative attacker intent and objectives models. Then model the attack strategies for each of the representative models. In Step 3, Step 4 and Step 5, do the same thing as the previous pipeline. As a result, a (separate) set of attack strategy inferences will be generated for each of the representative AIOS models built in Step 2.
- In the detection phase, once an attack strategy is observed, match the observed attack strategy against the inferred attack strategies generated in the learning phase. Once an inferred attack strategy is matched, the corresponding attacker intent and objective model(s) will be the inference(s) of the real attacker's intent and objectives. (Note that sometimes an observed attack strategy may "match" more than one attacker intent and objective models.) Nevertheless, when none of the inferred attack strategies can be matched, go back to the learning phase and do more learning.

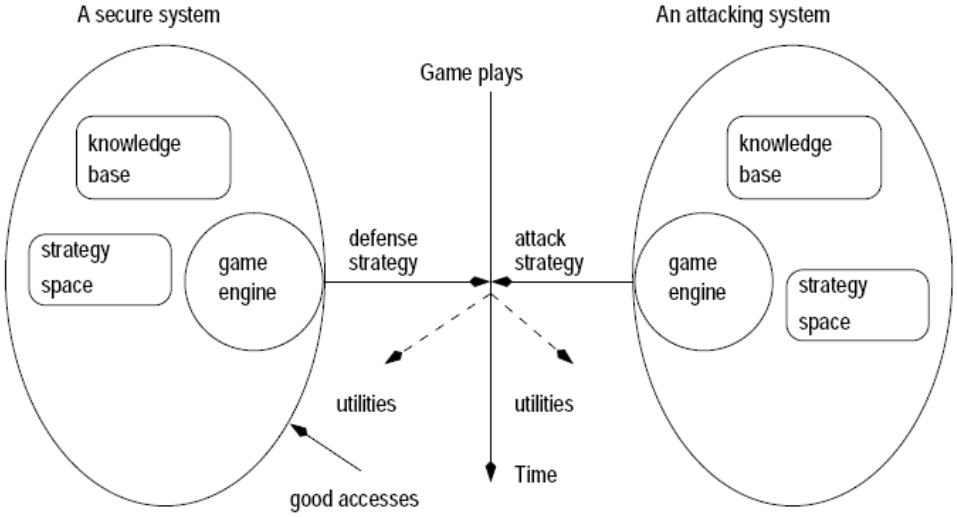



Figure 1. Game Theoretic Formalization of the Attack Prediction Framework

Within our framework, we also developed a game-theoretic AIOS inference model *taxonomy* which is shown in Figure 2. The taxonomy tells which type of game models (e.g., static, dynamic, Bayesian, stochastic) is good for which type of cyber systems. The taxonomy systematically addresses a set of important concepts, ideas and issues in predicting attacks:

- Using signaling games to predict attacks.
- Subgame-perfect Nash equilibrium strategies in attack prediction games.
- Static games vs. multi-stage dynamic games in predicting attacks.
- Applying the theory of *backwards induction* in predicting attacks.
- Bayesian repeated games vs. multi-stage dynamic games in predicting attacks.
- Predicting simultaneous attacks from multiple attacks.

Finally, to validate the hypothesis, we performed two specific case studies to show how attack strategies can be inferred in real world attack-defense scenarios. In particular, we developed a specific game theoretic prediction model for fraudulent credit card transactions, which we will describe shortly in Section C.2, and a specific prediction model for Internet distributed denial-of-service (DDoS) attacks, which we will describe shortly in Section C.3. We have done more than 5,000 simulation-based experiments to predict DDoS attacks and a family of intriguing and insightful predictive observations is obtained.

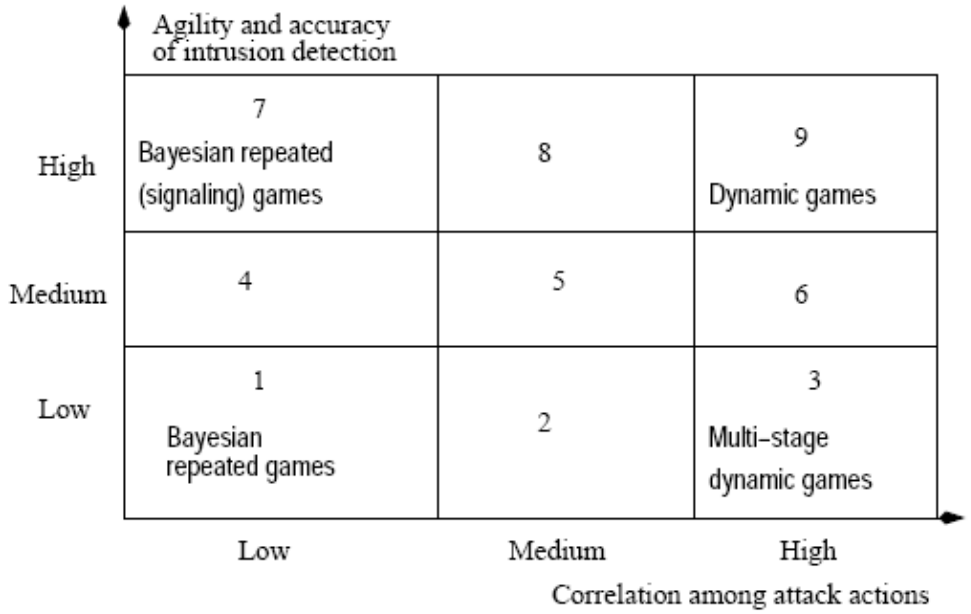



Figure 2. A Taxonomy of Game Theoretic AIOS Models

[**Impact**] We believe that this research is a major breakthrough in attack prediction, not only because the papers are accepted or selected by top security conferences and journals, but also because the developed framework for AIOS modeling and inference builds a scientifically sound foundation on top of which the merits and significance of game theoretic attack prediction models and methods can be well justified. In particular, the attack prediction methods we proposed in both **Task I** and **Task II** are all based on this framework, since accurate attack predictions against *intelligent* and *active* attackers are very difficult, if not impossible, to produce if the attacker's intent, objectives and strategies are neglected.

## C.2 Prediction, Detection and Analysis of Financial Cyber Crimes

As a major component of **Task I.A**, we developed a game theoretic approach to predict, detect and analyze financial cyber crimes such as credit card frauds and money laundering. The results are summarized in the following paper.

- Peng Liu, "Financial Cyber Crime Detection and Analysis: A Game Theoretic Approach", *International Journal of Information Policy, Law, and Security*, in review.

[**Motivation**] Financial cyber crimes (FCCs) are a serious threat to national security. Effective FCC defense (FCCD), especially FCC detection, can significantly improve out ability to

counter fraudulent financial transactions, criminal money laundering, and terrorism activities. However, existing FCC detection (and response) techniques have a fundamental limitation, that is, the way in which they determine the classification boundaries (e.g., FCCD thresholds) for the set of suspicion classes is ad hoc, cannot quantitatively bound the corresponding risk of national security loss, and may lead to disastrous consequences. This limitation is primarily due to the fact that existing FCC detection techniques cannot model or infer the intent, objectives and strategies of financial cyber criminals.

[**Hypothesis**] Our AIOS inference framework can be "materialized" or applied to predict financial cyber crimes (FCC) such as credit card frauds and money laundering.

[**Approach**] In the research, an innovative, game theoretic FCC detection optimization methodology is proposed, which is composed of a generic model of FCCD systems, an incentive-based conceptual model of FC criminal motives and strategies, and a game theoretic formalization which can capture the inherent inter-dependency between FC criminal motives and strategies and FCCD objectives and strategies in such a way that FC criminal motives and strategies can be predicted and optimal classification boundaries for FCC detection systems (FDS) can be estimated. This methodology is generic and can be used to build almost every type of FCC detection systems.

To validate the methodology, we have done a representative case study where the attacking strategies of credit card frauds are predicted. In this case study, we firstly built the game theoretic credit card fraud prediction model, which is shown in Figure 3, based on our AIOS inference framework.

(1) The game has two players: a subject $c_i$ and the FDS.

(2) $A_{fds}$ is the set of possible values for threshold $th(c_i)$. These values can be either continuous or discrete.

(3) $A_{c_i}$, the set of possible credit card transactions, is specified by the values of $amount(T_j^{c_i})$ for each transaction $T_j^{c_i}$.

(4) $T_{fds} = \{t^{fds}\}$; $T_{subject} = \{good, bad\}$. We assume at one point of time, there is only one fraud on $c_i$.

(5) We assume $p_{FDS}(good|t^{fds}) = 1 - \theta$, and $p_{FDS}(bad|t^{fds}) = \theta$.

(6) We assume $sl(T_j^{c_i}) = |amount(T_j^{c_i}) - P(c_i)|$.

(7) $$u_{subject}(th(c_i), T_j^{c_i}; good) = \begin{cases} 0 & \texttt{if } |amount(T_j^{c_i}) - P(c_i)| \leq th(c_i) \\ -DoS(T_j^{c_i}) & \texttt{if } |amount(T_j^{c_i}) - P(c_i)| > th(c_i) \end{cases}$$

(8) $$u_{subject}(th(c_i), T_j^{c_i}; bad) = \begin{cases} amount(T_j^{c_i}) & \texttt{if } |amount(T_j^{c_i}) - P(c_i)| \leq th(c_i) \\ 0 & \texttt{if } |amount(T_j^{c_i}) - P(c_i)| > th(c_i) \end{cases}$$

(9) $u_{fds} = (1-\theta)u_{fds}^{good}(th(c_i), T_j^{c_i}; t^{fds}) + \theta u_{fds}^{bad}(th(c_i), T_j^{c_i}; t^{fds})$, where

$$u_{fds}^{good}(th(c_i), T_j^{c_i}; t^{fds}) = \begin{cases} b.wsize & \texttt{if } |amount(T_j^{c_i}) - P(c_i)| \leq th(c_i) \\ 0 & \texttt{if } |amount(T_j^{c_i}) - P(c_i)| > th(c_i) \end{cases}$$

$$u_{fds}^{bad}(th(c_i), T_j^{c_i}; t^{fds}) = \begin{cases} -amount(T_j^{c_i}) & \texttt{if } |amount(T_j^{c_i}) - P(c_i)| \leq th(c_i) \\ 0 & \texttt{if } |amount(T_j^{c_i}) - P(c_i)| > th(c_i) \end{cases}$$

Figure 3. The game theoretic prediction model for credit card fraud

Secondly, we have developed a probabilistic numerical-analysis approach to solve the game formalized in Figure 3. We have also done extensive computer simulations on the game plays that capture the possible interactions among the fraud detection system, the credit card transactions issued by the real owner, and the credit card transactions issued by the fraud.

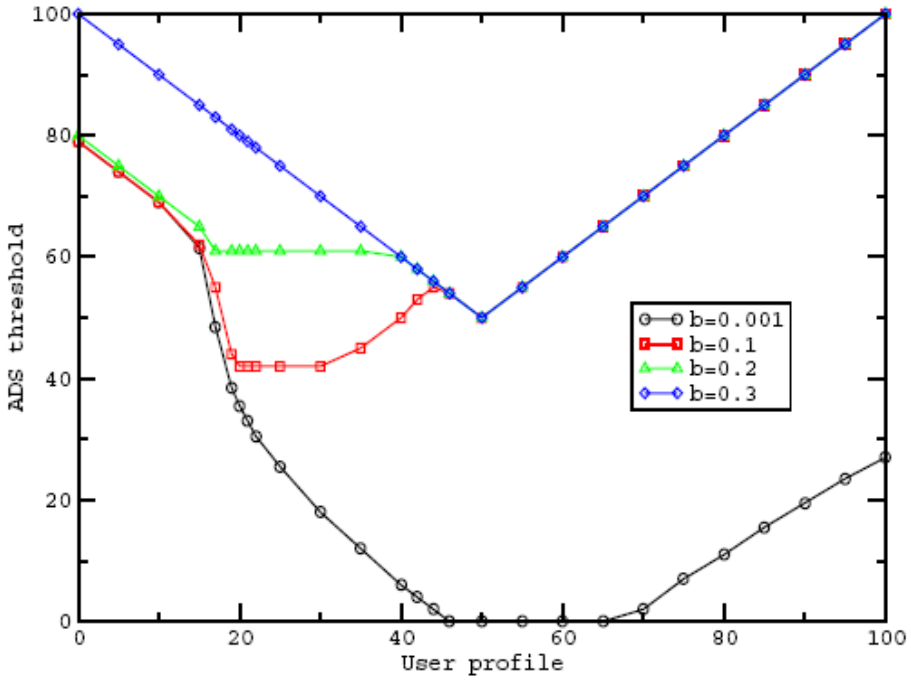



Figure 4. Nash equilibrium ADS strategies for each user profile when B=20 and θ is 0.05

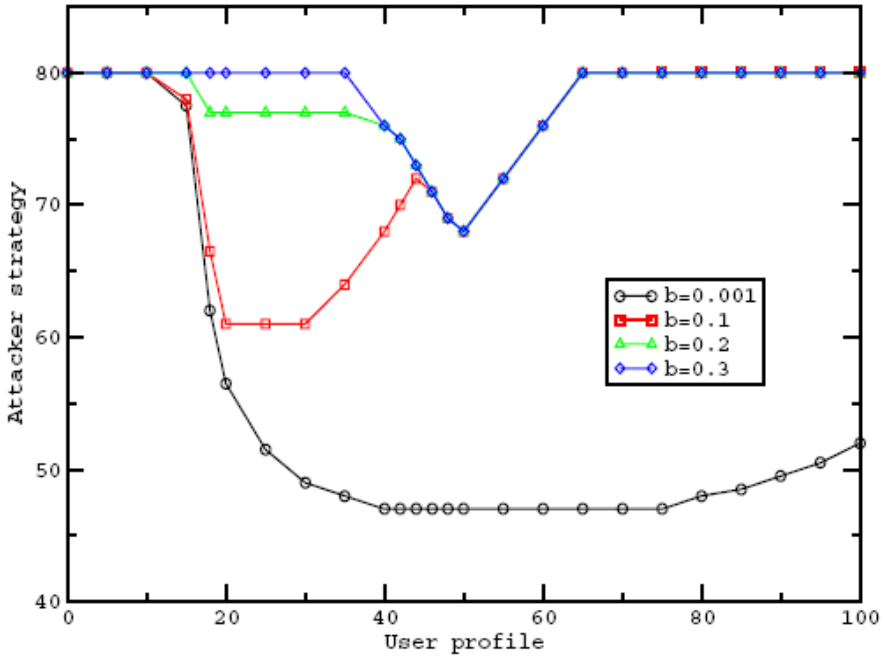



Figure 5. Nash equilibrium attacker strategies for each user profile when B=20 and θ is 0.05

Thirdly, we used the simulation results to calculate the Nash equilibria of the game theoretic interactions among the fraud detection system, the credit card transactions issued by the real owner, and the fraud transactions. Not surprisingly, we got the Nash equilibrium attacker strategies (as partially shown in Figure 4 and Figure 5) which are the 'predicted' attacker strategies that are most likely to be taken by rational credit card frauds.

Finally, we validated the attacker strategies predicted by the above approach via common sense analysis on credit card fraud and detection.

[**Impact**] We have found some interesting implications of these results.

1. The FDS should determine the value of the availability weight based on the credit card usage profile of the authentic customer.
2. The results show that (a) the cost of availability is higher attacker success rates and more security loss (i.e., attacker payoffs), and (b) the existence of Nash equilibria and optimal detection thresholds is due to the fact that the FDS needs to tradeoff between the goal of minimum fraud loss and goal of maximum service availability.
3. It should be noticed that the payoffs generated by the simulations are the expected payoffs of the attacker (FDS) instead of his or her *real* payoffs.
4. The Nash equilibria generated have interesting implications on the false alarm rate and the detection rate of the FDS.

## C.3 Prediction and Analysis of Internet DDoS Attacks

In **Task I.B**, we developed a game theoretic approach to predict the strategies that are most likely to be taken by Internet DDoS attackers. The results are summarized in the following paper.

- P. Liu, W. Zang, M. Yu, "Incentive-Based Modeling and Inference of Attacker Intent, Objectives and Strategies", *ACM Transactions on Information and Systems Security,* Vol. 8, No. 1, 2005, 78-118.

[**Motivation**] DDoS attacks are one of the most favorite attacks issued by attackers to cause serious effects in the Internet. The capability to predict DDoS attacking strategies may greatly help protect the Internet from DDoS attacks.

[**Hypothesis**] Our AIOS inference framework can be "materialized" or applied to predict Internet DDoS attacks.

[**Approach**] We developed a game theoretic approach to predict DDoS attacks against a network protected with the Pushback defense (developed by AT&T), and we have done extensive `ns2` simulation experiments (above 5000 experiments) to produce concrete DDoS attack predictions and justify the benefits of this approach.

In particular, (a) we built a multi-player Bayesian game model for the prediction purpose. The model consists of 10 components: the players (the attacker, the system, and several legitimate users), the attacker's action space, the legitimate users' action space, the system's action space, the attacker's type space, the system's type relief, the attacker and legitimate users' type belief, the attacker's utility, the legitimate users' utility, and the system's utility function.

(b) We have used the AT&T Pushback module to do above 5000 `ns2` simulation experiments to produce concrete DDoS attack predictions. Our year 1 simulations were based on the original Pushback topology composed of 64 hosts and 22 routers. To see whether the characteristics of the simulated attack/defense game plays and the corresponding conclusions we had drawn can still hold in a large scale DDoS attack/defense game, we have used *Brite*, a popular topology generator, to create a large network with 101 routers and more than 1000 hosts. And our experiment results show that (most of) the predictions we produced in year 1 still hold for large networks.

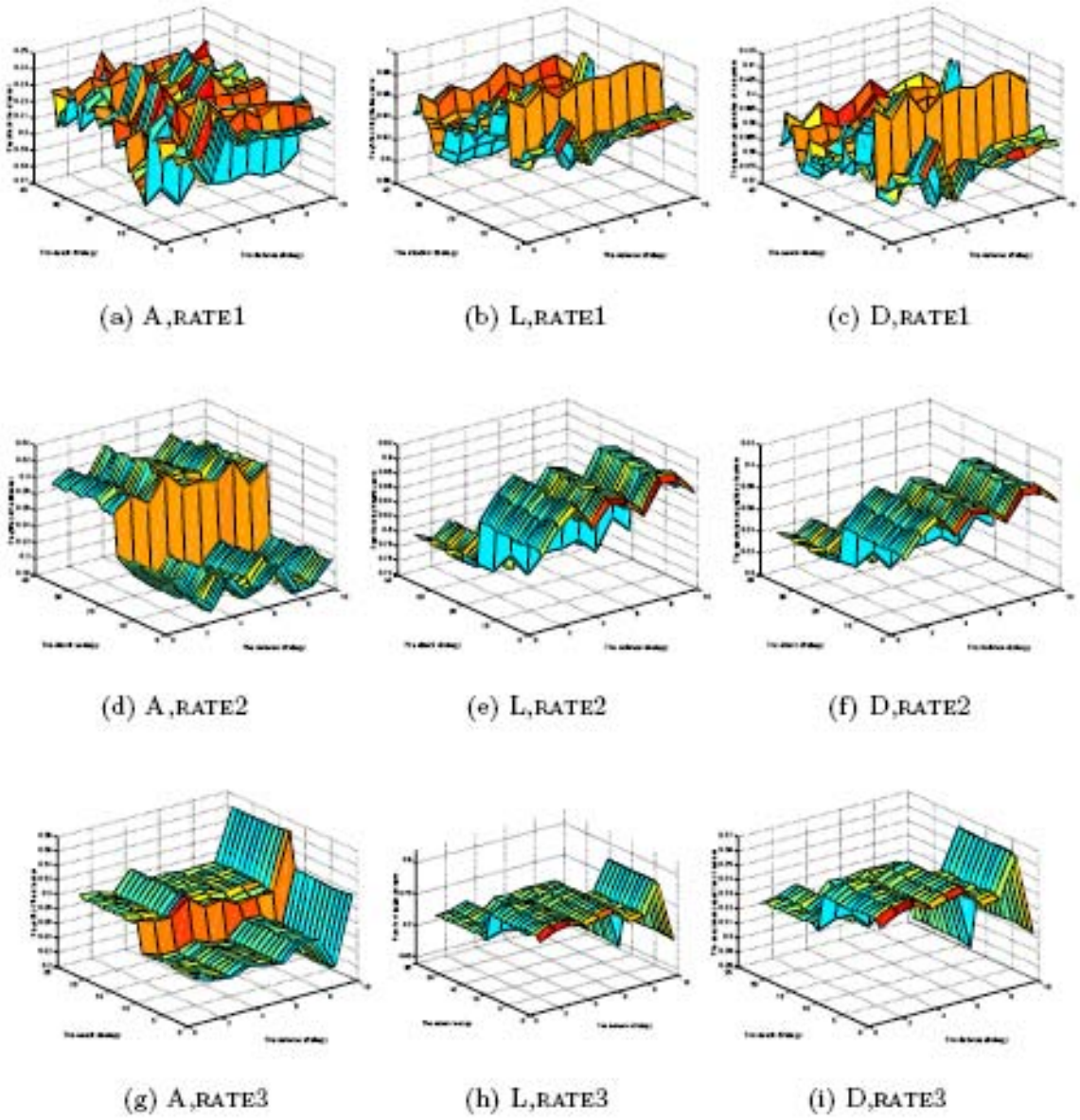



Figure 6. The Attacker's, Legitimate User's and Defense System's Payoffs under Different Defense and Attack Strategies

[**Findings**] First, based on the payoff results that are partially shown in Figure 6, we found that:

- The attacker's payoffs are dependent upon not only attack strategies, but also network scenarios and defense postures, which well justifies the Strategy Interdependency Property of our AIOS model.
- Our experiments confirm many well-known observations about DDoS attack and defense. For example, the attacker prefers more zombies and the defense system prefers lower drop rate. Nevertheless, our experiments give more insights on DDoS attack and defense. For example, many people believe that the attacker's and defense system's payoffs are mainly determined by the attack and defense strategies, but our results show that the ratio between the poor traffic volume and the total bandwidth is a very important factor, and this ratio may greatly affect the attacker's and defense system's payoffs.
- Our experiments also yield several surprising observations. (a) Many people may believe that the more packets the zombies send out to the victims, the more bandwidth and payoffs the attacker should earn. (b) Many people may believe that using different traffic patterns should be more effective in attacking than a single traffic pattern. (c) Many people may believe that exponential bit rate should be more effective in attacking than constant bit rate. (d) Many people may believe that using UDP should be more effective in attacking than TCP or ICMP. However, our results show that neither the attacking rate nor the traffic pattern matters, and different bad-to-poor ratios (30, 35 or 50) or different traffic patterns (UDP or ICMP) give the attacker similar amounts of payoffs.
- For the system, to obtain higher resilience against DDoS attacks, it need only be concerned with three specific pushback parameters, namely Target Drop Rate, Maximum Number of Sessions and Aggregate Pattern. The other parameters do not affect the results much.

| system strategy | Legitimate strategy | Attacking strategy |
|---|---|---|
| target-drop-rate = 0.03 | rate1, ManygoodFewpoor | Few, 35, CBR, One aggregate |
| target-drop-rate = 0.03 | rate1, ManygoodManypoor | Many, 30, CBR, One aggregates |
| target-drop-rate = 0.03 | rate1, ManygoodFewpoor | Many, 30, CBR, One aggregate |
| Maximum session = 5 | rate2, ManygoodManypoor | Many, 35, CBR, Multiple aggregates |
| Maximum session = 5 | rate2, ManygoodManypoor | Many, 30, EXP, Multiple aggregates |
| Maximum session = 5 | rate2, ManygoodManypoor | Many, 35, EXP, Multiple aggregates |
| Maximum session = 5 | rate2, ManygoodManypoor | Many, 45, EXP, Multiple aggregates |

Table 1. Nash Equilibrium Strategies

Second, based on our equilibrium results which are shown in Table 1, Table 2, Table 3 and Table 4, we found that:

- In terms of the traffic pattern, the distribution is shown in Table 2. "Dest" means the aggregate property is Destination Address Prefix. "DestPatt" means that the aggregate property is Destination Address Prefix plus Traffic Pattern. Table 2 shows that the attacker is more likely to use EXP traffic. In this way, he has more chances to stay at a Nash equilibrium since 50% Nash Equilibria occur when the traffic pattern is EXP. When

the aggregate is DestPatt, obviously, the attacker prefers to use the same traffic patterns as those used by poor and good users.

- The distribution under bad-to-poor ratio is shown in Table 3. Surprisingly, the distribution shows that the attacker is most unlikely to use a high ratio. Some people may believe that higher bad-to-poor ratio should make the attack more successful since more packets will be flooded to the victim(s). However, our analysis of Nash Equilibria distributions shows that the attacker has better opportunities to converge to a Nash equilibrium strategy with low bad-to-poor ratio. We believe that an important reason for this phenomenon is because our DDoS game is not a zero-sum game.
- The distribution under different combinations of the number of zombies, poor hosts and good hosts is shown in Table 4. In the table, "F" means "Few" and "M" means "Many". "FMF" means "FewgoodManypoorFewbad". The table indicates that the attacker prefers to use as many zombies as possible, which is consistent with the common sense of DDoS attacks.
- The distribution under different defense strategies indicates that most Nash equilibria occur when the target-drop-rate is 0.03 or when the max-number-of-sessions is 5. The probability that Nash equilibria occur under target-drop-rate 0.03 is 0.45, and under max-number-of-session 5 is 0.36. Hence, to be more resilient, the system can increase the number of sessions and decrease the target-drop-rate. Our analysis also shows that the impact of other defense strategy parameters on this distribution is minimum.

| Aggregate property | CBR | EXP | ICMP | MIXED |
|---|---|---|---|---|
| Dest | 0.09 | 0.50 | 0.27 | 0.14 |
| DestPatt | 0.38 | 0.25 | 0 | 0.38 |

Table 2. Nash Equilibrium Distribution under Different Attacking Patterns

| Aggregate property | 30 | 35 | 40 | 45 | 50 |
|---|---|---|---|---|---|
| Dest | 0.23 | 0.32 | 0.14 | 0.09 | 0.23 |
| DestPatt | 0.50 | 0 | 0.13 | 0.25 | 0.13 |

Table 3. Nash Equilibrium Distribution under Different Attacking Ratios

| Aggregate property | FFF | FFM | FMF | FMM | MFF | MFM | MMF | MMM |
|---|---|---|---|---|---|---|---|---|
| Dest | 0 | 0 | 0 | 0 | 0 | 0.09 | 0.55 | 0.36 |
| DestPatt | 0.12 | 0 | 0.13 | 0.13 | 0 | 0 | 0.38 | 0.25 |

Table 4. Nash Equilibrium Distribution under Different Number of Users

## C.4 Predicting the Resilience of Computer Systems against Attacks

We have adjusted the objective of **Task I.C** to predict the resilience of computer systems against cyber attacks. In particular, we developed an Economics Theoretic framework for Measuring Assurance, called EMTA, and measured the Internet's resilience against DDoS attacks. The main results are summarized in the following paper.

- W. Zang, P. Liu, M. Yu, "A Game Theoretic Analysis of Resilience of Internet against DDoS Attacks", to be submitted for journal publication.

[**Motivation**] Current information assurance techniques do not allow us to state quantitatively how resilient or assured our systems and networks are. Without quantitative statements about assurance, (a) people cannot have a tangible understanding about how assured our systems and networks are. (b) It is difficult to characterize the capabilities of protective, detection, reactive, proactive or self-regenerative security measures, such as firewalls, intrusion detection systems, self-healing techniques, design diverse redundancy, proactive secret sharing, and deception. (c) It is difficult for people to compare the capabilities of two security measures; it is difficult for people to compare the assurance of two secure information systems. (d) People cannot find a tangible correlation between a qualitative security evaluation statement and the amount of assurance they actually get. (e) Security and assurance measures can only be designed in an ad hoc fashion, based solely on what feels right, as opposed to whether the design can meet a quantitatively stated assurance requirement. (f) Security and assurance can only be built into information systems in an ad hoc fashion, based solely on what can be afforded or what feels right, as opposed to what is desired or required for a given application and its operating environment. There is no guarantee that systems designed as such will be effectively protected when under a sustained cyber attack. The key to solve the above problems is the idea of "Measuring Assurance in Cyber Space", where measures of merit and metrics to characterize quantitatively various dimensions of security (availability, integrity, confidentiality, authentication, and non-repudiation) are identified, modeled, measured, monitored, evaluated, and controlled. If this idea bears fruit, researchers and designers of information systems security will be able to make quantitative evaluations of novel architectural approaches, perform cost-benefit trade-offs, and create designs that meet specified levels of assurance.

[**Hypothesis**] Our AIOS framework can be extended to build a game theoretic framework to measure the resilience of computer systems against attacks.

[**Approach**] We proposed the EMTA framework. The key idea is using incentive-based, economic models of attacker intent, objectives, and strategies (AIOS) to measure a system's (overall) resilient or assurance-capacity. Compared with reliability measuring, a unique challenge to assurance measuring is that attacks are not *random*. As a result, the combination of a well-defined system, a complete threat, vulnerability, attack, and risk (TVAR) taxonomy (in terms of the system), a representative set of 5 assurance metrics, a representative workload, and (even) a rich attack history on the system may not be enough to yield accurate measurements about the system's (overall) assurance capacity, since statistics about old attacks may not capture the characteristics of new attacks which are intentional and not random.

Measuring a system's assurance capacity needs the ability to measure the attacker's attack capacity, which is however inter-dependent on the system's defense capacity. Hence, measuring the attacker's attack capacity needs the ability to model the attacker's IOS, which is however interdependent on the system's IOS.

An incentive-based, economic model of AIOS not only models the attacker's IOS but also models the system's IOS. Moreover, this new model uses game theory to model

mathematically the inter-dependency between these two IOS. And such a mathematical model can generate valuable, quantitative inferences about both the attacker's attack capacity and the system's defense capacity. These inferences can then be used to generate valuable quantitative measurements about the system's assurance capacity. Note that this model seamlessly integrates system specifications, TVAR, assurance metrics, workload, and attacks.

In particular, the economics theoretic assurance measuring framework is composed of the following components. Note that they are related to each other. The modeling language is used to specify the family of cyber security models. The AIOS model is built on top of the family of cyber security models. The AIOS model uses IA metrics and the corresponding measurements to define the utility earn by either the attacker or the system. The assurance measurements generated by our framework is based on the minimum utilities that could be earned by the system when a specific set of equilibrium defense strategies are taken by the system. The set of equilibrium defense strategies are determined based on the AIOS model. AIOS inferences can be valuable attack action predictions. The assurance measurements generated by our framework provide a lot of useful hints for optimal security design, and such measurements are natural means of security evaluation.

- An expressive modeling language that can express a variety of IA domain issues.
- A family of novel cyber security models that model cyber security not only from the defense perspective, but also from the attack (or offense) perspective. To our best knowledge, this family is the first security model that can model how the attacker and the de6 fender (i.e., the computer system) can interact dynamically on each other.
- An incentive-based, economic model of AIOS. This model can compute valuable AIOS inferences.
- An incentive-based, economic interpretation of resilience metrics and measurements that are security mechanism independent.
- An economic, game-theoretic approach of resilience measuring, namely, using economic utilities and equilibrium AIOS inferences to measure a system's overall assurance capacity.
- An economic, game-theoretic approach to predict attack actions with confidence.
- An economic, game-theoretic method of optimal security design. This method can identify the key design issues and factors, and can evaluate the benefits of a new security design either comparatively or on an absolute scale.
- An economic, game-theoretic method of security evaluation. While existing security evaluation techniques are primarily qualitative, this method is quantitative.

[**More Details**] The input-output semantics of the ETMA framework is shown in Figure 7. The inputs of the ETMA framework cover every (important) element of information assurance, such as the system (including the set of protection measures), the attacker, the environment (e.g., the workload), vulnerabilities, threats, risks, attacks, attack effects, defense postures, and defense actions. On the other hand, the outputs of the ETMA framework not only give us quantitative assurance measurements, which are the primary goal of this seedling effort, but also give us a new security modeling methodology, a new security design methodology, a new attacker modeling methodology (i.e., AIOS modeling and inferring), a

new attack prediction methodology, and a new security evaluation methodology at the same time.

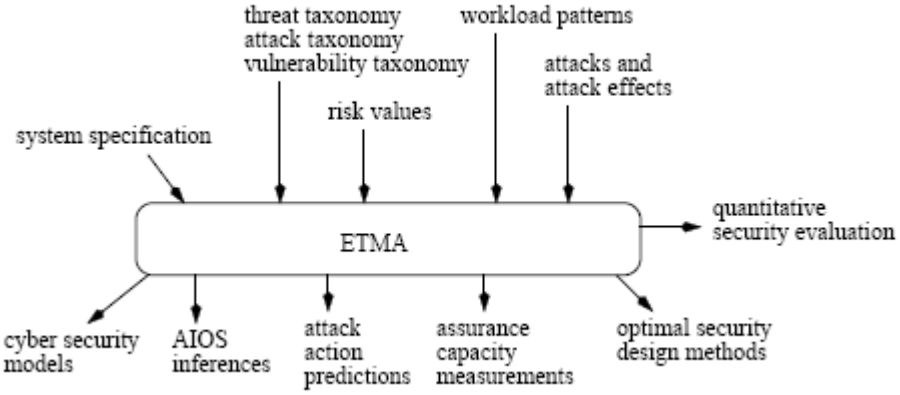



Figure 7: The Input-Output Semantics of the ETMA Framework

[**Case Study**] To evaluate the ETMA framework, we have done a case study where the Internet's resilience against DDoS attacks is measured. The experiments we have done in this case study are roughly the same as the experiments done in Section C.3, but the goal here is to infer the defense capacity of a computer system (e.g., a network) and the attack capacity of an attacker, instead of predicting the attacking strategies that are most likely to be taken by the attacker. Accordingly, we need to measure some different metrics.

Besides the experiment results shown in Figure 6 and Tables 1-4, we generated a set of results on the defense capacity of a computer system and the attack capacity of an attacker, as shown in Table 5. These results show that:

- Based on the set of Nash equilibria calculated, we can get the upper bounds of the attacking capacity of the attacker and the upper bounds of the resilience or assurance-capacity of the defense system under different network scenarios. These upper bounds are shown in Table 5. In particular, the upper bounds of the assurance-capacity tell us how well the system (i.e., Pushback) is resilient to DDoS attacks. The upper bounds of the attacking capacity tell us how serious the damage could be in the worst case. According to the definition of payoff function, the highest attacking capacity is 1 and the highest defense capacity is $\theta$, which is 0.5 in the paper. We used different normal traffic towards the victim to represent different network scenarios in the paper. When the traffic rate is very low, such as `rate1`, no matter how hard the attacker tries, the highest attacking capacity he could get is only 0.2076. And the highest assurance capacity of the defense system is 0.3941. When the traffic rate is high, such as `rate3`, the highest attacking capacity is 0.2862, which substantial higher.

| Network scenario | Attacking capacity | Assurance capacity |
|---|---|---|
| rate1 | 0.2076 | 0.3941 |
| rate2 | 0.2668 | 0.3820 |
| rate3 | 0.2862 | 0.3320 |

Table 5. Upper Bounds of the Assurance Capacity and Attacking Capacity

[**Impact**] We believe the impact of the research on measuring resilience will go beyond measuring. First, to quantitatively measure resilience, we need to quantitatively model secure information systems and the relevant IA domain issues. Hence the research on measuring resilience can motivate new cyber security models. Second, to measure an information system's resilience against intentional, well-planned attackers who issue non-random attacks, we need to model attacker intent, objectives, and strategies. Hence the research on measuring resilience can motivate new attacker models. Third, since quantitative assurance measurements can give people great leverage in designing better secure systems, the research on measuring resilience may revolutionize security design methodologies. Finally, the research on measuring resilience may revolutionize the way people evaluate security and assurance, the way vendors promote their products, and the way people deploy security and assurance measures.

## C.5 Attack Resistant Workflow Systems

To successfully perform **Task II.A**, namely predicting unknown or new types of data corruption attacks, we need to first investigate (predictive) defenses against data (and code) corruption attacks in such information systems as workflow systems before investigating how to build game theoretic prediction models against these unknown or new attacks. The rational is that good attack prediction models cannot be developed without a good understanding of both the attacks and the corresponding defense mechanisms.

Accordingly, we have developed an attack resistant, self-healing workflow system. Some results of this research are summarized in the following papers.

- M. Yu, W. Zang, P. Liu, "Defensive Execution of Transactional Processes against Attacks", In *Proc.* ACSAC '05, 2005, To appear, Acceptance rate = 19.6%
- M. Yu, P. Liu, W. Zang, "Self Healing Workflow Systems under Attacks", in *Proc. 24th IEEE International Conference on Distributed Computing Systems* (ICDCS 04), Tokyo, Japan, March 2004, pages 418-425. Acceptance rate = 17.68%

[**Motivation**] Increasingly, workflow management systems become the primary technology for organizations to perform their daily business processes in various important applications such as financial services and infrastructure management (e.g. transportation scheduling). Unfortunately, inside or outside cyber attacks against the applications can subtly manipulate data or processes in such a way that detection is hard but the damage/taint may widely spread throughout the system via the attacks' domino-effects.

Because the domino-effects can lead to disastrous real world effects, attack resilience and recovery are necessary. However, existing workflow recovery techniques require stop of services for repair, which can cause unacceptable denial-of-service. Our goal is to develop an attack resilient, self-healing workflow system such that for any real world info processing

application that involves distributed business processes, this technology can help automatically remove the destructive effects of cyber attacks on critical data and processes in real time without stopping good services.

[**Originality**] Compared with the state of the art, (a) current approaches use checkpoints, our approach does dependency-based taint tracing; (b) current approaches need to stop the services during repair, our approach does repair on-the-fly; (c) current approaches abort both good and tainted work, our approach keeps the untainted work; (d) current approaches are reactive, our approach can do both reactive self-healing and proactive self-healing via our "defensive execution" technology.

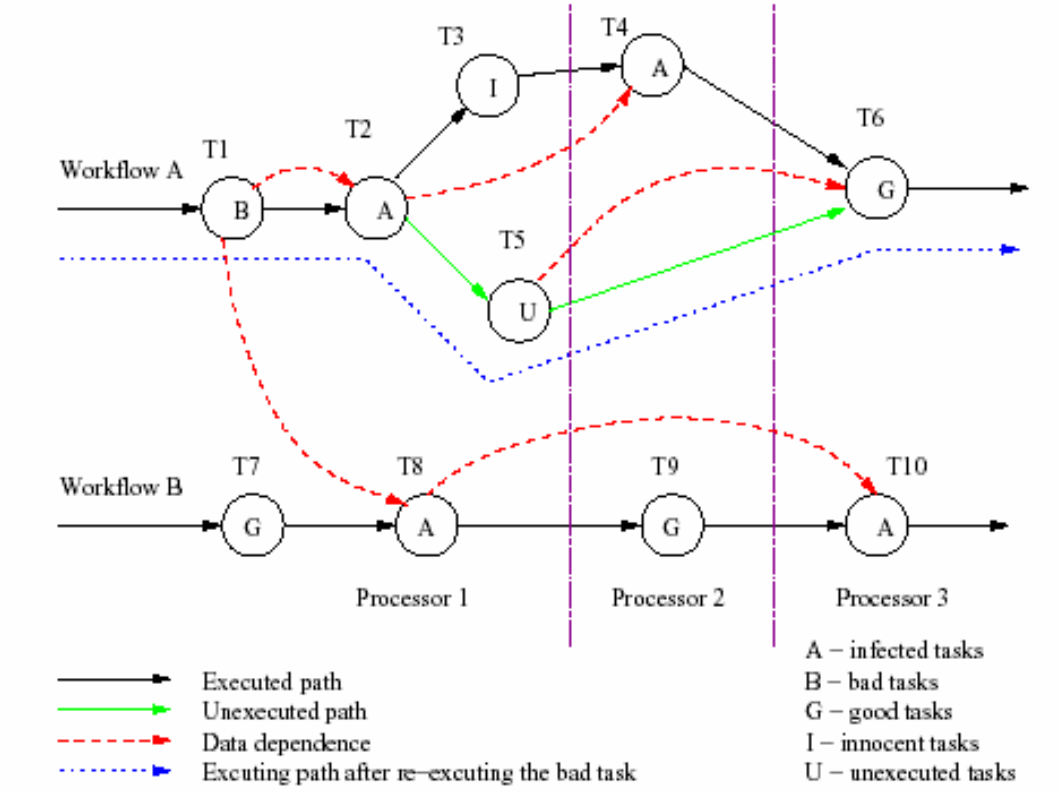
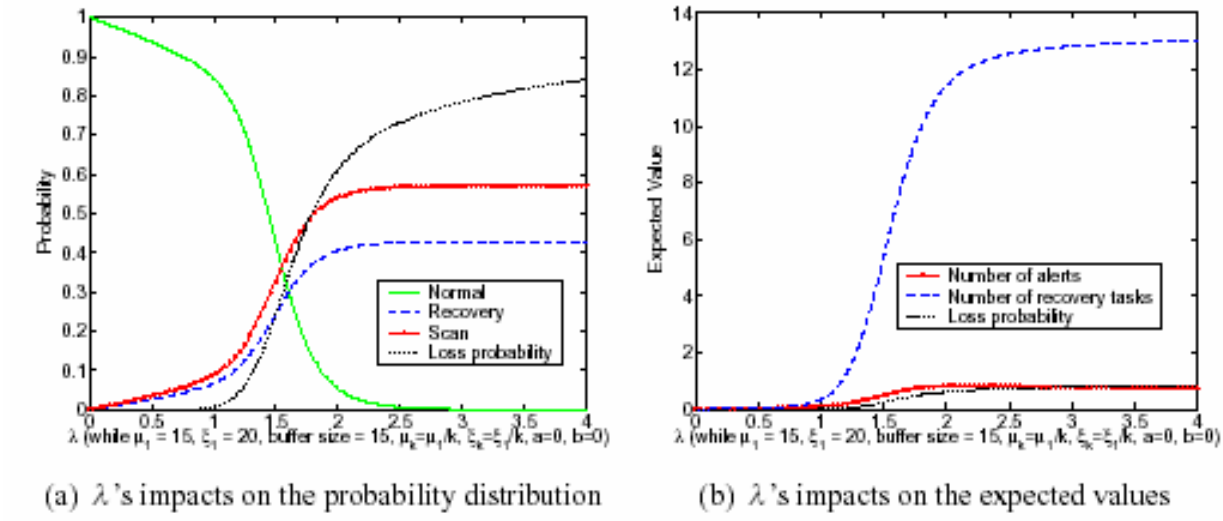



Figure 8. Two distributed business processes

[**Approach**] A workflow is a business process that consists of a partial order of tasks. For example, two workflows are shown in Figure 8. A task is "tainted" if it reads a tainted data object or is control dependent on another tainted task. A data object is tainted if it is updated by a tainted task or an attack. To guarantee correctness of recovery, we need to use three types of data dependencies (i.e., *flow-dependency*, *anti-flow dependency*, *output dependency*) and the must-be-executed-before *control dependency* among tasks. Accordingly, we have developed the first on-the-fly attack recovery theory for workflow systems in the face of cyber attacks.

Next, we have built a self-healing workflow system prototype based on the attack recovery theory and developed all the corresponding workflow self-healing algorithms. In our system, the recovery process has three possible modes: the "normal" mode when no attack is detected; the "scan" mode when the log is analyzed to locate undo and redo tasks; and the "repair" mode, where undo, redo and normal tasks must be carefully ordered based on both data and control dependencies.

To evaluate the cost-effectiveness of the proposed self-healing workflow technology, we have done both mathematical analysis and experimental performance evaluation using the prototype under different attacking densities, intrusion detection delays and arrival rates. The performance analysis is done by modeling the mode transitions as a Continuous Time Markov Chain, and some analysis results are shown in Figure 9. Our analysis and experimental evaluation results demonstrate that our self-healing workflow system is attack resilient and practical when parameters of the system are reasonably set up.




(a) $\lambda$'s impacts on the probability distribution (b) $\lambda$'s impacts on the expected values

Figure 9. Some Performance Analysis Results of Workflow Self-Healing

## C.6 Other Partially Supported Researches

Finally, several other relevant researches we have been doing on such topics as secure information sharing (in collaborative information processing) and intrusion masking distributed computing are also partially supported by this grant.

# D. Products Developed under the Award

## D.1 Publications

1. P. Liu, W. Zang, M. Yu, "Incentive-Based Modeling and Inference of Attacker Intent, Objectives and Strategies", *ACM Transactions on Information and Systems Security,* 56(3): 283—298.
2. Peng Liu, "Financial Cyber Crime Detection and Analysis: A Game Theoretic Approach", *International Journal of Information Policy, Law, and Security*, under review.
3. P. Liu, W. Zang, "Incentive-Based Modeling and Inference of Attacker Intent, Objectives and Strategies," *Proc. 10th ACM Conference on Computer and Communications Security (CCS '03)*, October 28-31, Washington DC, 2003, pages 179-189.
4. W. Zang, P. Liu, M. Yu, "A Game Theoretic Analysis of Resilience of Internet against DDoS Attacks", to be submitted for journal publication.
5. Peng Liu, Meng Yu, Jiwu Jing, "Information Assurance", In *Handbook of Information Security*, Hossein Bidgoli et al. (eds.), John Wiley & Sons, 2005.
6. M. Yu, P. Liu, W. Zang, "Self Healing Workflow Systems under Attacks", in *Proc. 24th IEEE International Conference on Distributed Computing Systems* (ICDCS 04), Tokyo, Japan, March 2004, pages 418-425. Acceptance rate = 17.68%
7. M. Yu, W. Zang, P. Liu, "Defensive Execution of Transactional Processes against Attacks", In *Proc.* ACSAC '05, 2005, to appear, Acceptance rate = 19.6%
8. M. Yu, P. Liu, W. Zang, "Dependency Relation based Attack Recovery of Workflow Systems", *ACM Transactions on Information and Systems Security*, in review.
9. M. Yu, P. Liu, W. Zang, "The Implementation and Evaluation of a Self-Healing Workflow System," *IEEE Transactions on Dependable and Secure Computing*, in review for journal publication.
10. M. Yu, P. Liu, W. Zang, "Specifying and Using Group-to-Group Communication Services for Intrusion Masking", *Journal of Computer Security,* Vol. 13, No. 4, 623-658.
11. H. Wang, P. Liu, L. Li, "Evaluating the Impact of Intrusion Detection Deficiencies on the Cost-Effectiveness of Attack Recovery, *Proceedings of the 7th Information Security Conference*, San Francisco, September 2004.
12. R. Li, J. Li, H. Kameda, P. Liu, "Localized Public-key Management for Mobile Ad Hoc Networks", *Proc. 2004 IEEE Globecom*, Nov 2004.
13. Q. Gu, P. Liu, W. Lee, C. Chu, "eKTR: An Efficient Key Management Scheme in Wireless Data Broadcast Services", *Proc. 2005 IEEE Mobiquitous*, short paper, 2005.
14. Peng Liu, Hai Wang, Lunquan Li, "Real-Time Data Attack Isolation for Commercial Database Applications", *Elsevier Journal of Network and Computer Applications*, in press.
15. Q. Gu, P. Liu, S. Zhu, C. Chu, "Defending against Packet Injection Attacks in Unreliable Ad Hoc Networks", In *Proc. IEEE GLOBECOM '05*, 2005, to appear.
16. Peng Liu, Amit Chetal, "Trust-Based Secure Info Sharing Between Federal Government Agencies", *Journal of the American Society for Information Science and Technology*, Vol. 56, No. 3, 2005, pages 283-298.
17. R. Li, J. Li, P. Liu, H. H. Chen, "On-Demand Public-Key Management for Mobile Ad Hoc Networks", *Journal of Wireless Communications and Mobile Computing*, accepted, to appear.

18. Q. Gu, P. Liu, C. Chu, "Analysis of Area-congestion-based DDoS Attacks in Ad Hoc Networks", *International Journal of Ad Hoc Networks*, in review.

## D.2 Web Sites

The results of this project have been reflected at the following Web sites:

http://ist.psu.edu/s2/Prediction.html

http://ist.psu.edu/s2/research.html

http://ist.psu.edu/s2/

## D.3 Other Products

- [Data Set] We generated a large data set of DDoS attack simulations against Pushback based defense. Each entry of the data set describes the simulation results of a unique set of DDoS attacking parameters and defense parameters.
- [Software] We developed and implemented a self-healing workflow system prototype.

## E. Education and Academic Training

- Number of post docs: 2 (Dr. Wanyu Zang (female) and Dr. Meng Yu)
- Number of graduate students: 4 (Qijun Gu, Lunquan Li, Kun Bai, Hai Wang)
- Number of undergraduates students: 3 (Dan Gao (female), Charles Nwatu, Joshua Blackman)
- This DOE project is now being integrated into the capacity effort associated with the PSU Center for Information Assurance, a National *Center of Excellence* in IA Education designated by NSA.

## Acknowledgement

This material is based upon work supported by the U.S. Department of Energy under Award DE-FG02-02ER25527.