Near Optimal Attack Strategies for the Maximization of Information Theft

Dr. Frank Y.S. LIN
Department of Information Management, National Taiwan University
No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan

Chung-lien TSENG[+]
Department of Information Management, National Taiwan University
No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan

Po-Hao TSANG
Department of Information Management, National Taiwan University
No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan

Abstract

With the prevalence and varied applications of the Internet, new cyber-crimes are mushrooming all over cyberspace. The crimes are characterized by their "silent" attack behavior, which enables an attacker to exploit the vulnerabilities of a system and steal information, without actually crashing the system. Information theft is a relatively new cyber-crime that not only causes property damage and monetary loss to its victims, but can also ruin their reputations.

To detect and analyze the serious impact of information theft, we model it as a mathematical programming problem, defined by the AS model. In the model, an attacker applies his limited attack power intelligently to the targeted network in order to steal as much valuable information as possible. A Lagrangean relaxation-based algorithm is adopted to solve the AS problem, and the "susceptibility" metric is used to evaluate the effect of the attack.

Keywords: Information Theft, Lagrangean Relaxation, Network Attack, Optimization Problem, Resource Allocation, Scale-free Networks.

1. Introduction

With the prevalence and varied applications of the Internet, new cyber-crimes are mushrooming all over cyberspace. Unlike attackers in the past, who tried to crash a whole network or interrupt a system's normal services, attackers now tend to exploit the vulnerabilities of a system and steal information, without actually crashing the system. Information theft is a relatively new cyber-crime characterized by this "silent" attack behavior. It not only causes property damage and monetary loss to its victims, but can also ruin their reputations.

Since an attack does not affect normal network operations, an occurrence can easily be missed. Usually, it is too late when the victim realizes that a network or system has been compromised because the damage has been done. To prevent such occurrences, network operators can invest some resources to enhance the robustness of the whole network. However, since resources are limited, it is impossible to make a network entirely attack-proof; thus, a network operator must allocate his limited resources effectively.

Before determining the best defense resource allocation strategy, we must first consider the best attack strategy. This is a case of "know your enemy and know yourself." Previous research has shown that attempts to model attackers' actions in an abstract, mathematical way and then predict the attackers' future tactics based on those models is a non-trivial and unsolved issue [1, 2]. Therefore, in this paper, we model the attacker's behavior as a mathematical formulation, and compare the robustness of different network topologies under different defense budget allocation strategies against malicious attacks.

2. Attack Scenario and Problem Formulation

2.1. Problem Description

Because an attacker's resources, i.e., time and money, are limited, only part of a network can be compromised.

[+] Correspondence should be sent to r93002@im.ntu.edu.tw

Therefore, the resources must be fully utilized so that the attacker can gain the most valuable information that will cause the maximum harm to the network operator.

Of course, the reward an attacker can gain may change when the defense resource allocation strategy changes. Hence, to evaluate the efficiency of an attack under different defense strategies, we analyze the susceptibility of the network. The susceptibility metric, shown in the Equation 1, is defined as the percentage of stolen information. It is assumed that the attacker can steal all the information held by a node once the node is compromised successfully. Assume that d_i is the value of information held by node i, where $i \in N$.

$$Susceptibility(\%) = \left(\frac{\sum_{i \in nodes\ that\ are\ compromised} d_i}{\sum_{j \in all\ nodes\ in\ the\ network} d_j}\right) \times 100\% \qquad (1)$$

Note that the network we discuss here is at the AS level.

2.2. Problem Formulation of the AS Model

The attack scenario is as follows. Initially, the attacker controls one node that connects directly to the targeted network, and that node is viewed as the initial hop-site to reach other nodes. Since the targeted network is at the AS level, the attacker cannot simply attack any node directly. Instead, he can only reach uncompromised nodes from their immediate compromised neighbors. Thus, the attacker needs to construct an *attack tree*, i.e., a tree consisting of compromised nodes and rooted at the initial hop-site. To consider the worst case scenario, we assume the attacker is smart enough to obtain complete information about the targeted network in advance.

The effort needed to compromise a node depends on the resources allocated to defend the node. Generally, the more defense resources a node has, the more robust it is. However, a node still has some defense capability, even if no defense resources are allotted to it, since the node itself is a shell for protecting the information. On the other hand, the total attack resources are limited by the allocated budget. Our objective is to understand how an attacker can distribute his limited resources effectively and intelligently in order to maximize his benefit. To achieve our objective, we formulate the above problem as a maximization mathematical model (AS model).

Table 2-1 Given parameters of the AS model

Notion	Description
N	The index set of all nodes in the network
W	The set of all O-D pairs, where the origin is node s; and the destinations are the nodes with positive d_i , where $i, s \in N$
d_i	Damage incurred by compromising node i, where $i \in N$
P_w	The index set of all candidate paths of an O-D pair w, where $w \in W$
A	The total attack power
$\hat{a}_i(b_i)$	The threshold of the attack power required to compromise node i, i.e., the defense capability of node i, where $i \in N$
δ_{pi}	An indicator function, which is 1 if node i is on path p; and 0 otherwise (where $i \in N$, $p \in P_w$)

Table 2-2 Decision variables of the AS model

Notion	Description
a_i	Attack power applied to node i, where $i \in N$
y_i	1 if node i is compromised; and 0 otherwise (where $i \in N$)
x_p	1 if path p is selected as the attack path; and 0 otherwise (where $p \in P_w$)

Objective function

$$Z_{\text{IP}} = \max_{y_i, a_i} \sum_{i \in N} d_i y_i \equiv \min_{y_i, a_i} -\sum_{i \in N} d_i y_i \text{ ,} \qquad \textbf{(IP 1)}$$

subject to:

$$\sum_{w \in W} \sum_{p \in P_w} x_p \delta_{pi} \le (|N| - 1) y_i \qquad \forall\, i \in N \qquad \text{(IP 1.2)}$$

$$\sum_{p \in P_w} x_p = y_i \qquad \forall\, i \in N,\ w = (s, i) \qquad \text{(IP 1.3)}$$

$$\sum_{p \in P_w} x_p \le 1 \qquad \forall\, w \in W \qquad \text{(IP 1.4)}$$

$$x_p = 0 \text{ } or \text{ } 1 \qquad \forall\, p \in P_w,\ w \in W \qquad \text{(IP 1.5)}$$

$$y_i = 0 \text{ } or \text{ } 1 \qquad \forall\, i \in N \qquad \text{(IP 1.6)}$$

$$0 \le a_i \le \hat{a}_i(b_i) \qquad \forall\, i \in N \qquad \text{(IP 1.7)}$$

$$\sum_{i \in N} a_i \le A \qquad \text{(IP 1.8)}$$

$$\hat{a}_i(b_i) y_i \le a_i \qquad \forall\, i \in N\,. \qquad \text{(IP 1.9)}$$

The objective of the formulation is to maximize the total value of the information stolen. Constraints (IP 1.1) ~ (IP 1.5) jointly require that, when a node is chosen for attack, there must be exactly one path from the attacker's initial position, s, to that node, and each node on the path must have been compromised. These constraints are jointly described as the "continuity constraints." The above formulation can be viewed as a 0-1 knapsack problem with continuity constraints, where each node represents an item, and the node's information value and defense capability are the item's profit and weight respectively.

3. Solution Approach

3.1. Lagrangean Relaxation-based Algorithm

We propose a Lagrangean relaxation-based algorithm [3], which we denote as LR, in conjunction with the subgradient method [3] to solve the AS model. To achieve better results, a two-stage Lagrangean relaxation procedure is adopted. In the first stage, we relax Constraints (IP 1.1), (IP 1.2), and (IP 1.8), and construct a Lagrangean relaxation problem (LR 1). In the second stage, (IP 1) is transformed into another Lagrangean relaxation problem (LR 2) by relaxing Constraints (IP 1.1), (IP 1.2), and (IP 1.7).

The relaxed problems are then solved optimally to get a lower bound for the primal problem. After solving (LR 1), the resulting bounds are taken as the initial bounds in the second stage. Two heuristics are adopted to derive feasible solutions to the primal problem, and the subgradient method is used to update the Lagrangean multipliers. The time complexity of each iteration in the LR procedure is $O(|N|\log^2|N|)$.

3.2. First-Stage Relaxation

Lagrangean relaxation problem

$$Z_{\mathrm{D}}(\mu_1, \mu_2, \mu_3) = \min_{y_i} \quad \textbf{(LR 1)}$$

$$-\sum_{i\in N} d_i y_i + \sum_{i\in N} \mu_i^1 [\sum_{w\in W} \sum_{p\in P_w} x_p \delta_{pi} - (|N|-1) y_i] + \sum_{i\in N} \mu_i^2 [\sum_{p\in P_{(s,i)}} x_p - y_i] + \sum_{i\in N} \mu_i^3 [\hat{a}_i(b_i) y_i - a_i],$$

subject to:
(IP 1.3) ~ (IP 1.7), and

$$u_i^1, u_i^3 \geq 0. \qquad \forall\, i \in N. \qquad \text{(LR 1.1)}$$

We decompose (LR 1) into three independent and easily solvable optimization subproblems with respect to decision variables x_p, y_i, and a_i, and solve the respective subproblems optimally.

Getting primal feasible solutions

Solutions to (LR 1) and their associated Lagrangean multipliers are considered in order to obtain a primal feasible solution for (IP 1). The concept of the proposed heuristic, denoted as Heuristic_LR_1, is described below.

The main concept of this greedy-based heuristic arises from the attacker's strategy of compromising nodes with smaller weights but moderate path costs in order to maximize the gain. Thus, only attack paths comprised of activated nodes, i.e., nodes with smaller weights, will be constructed. The total computational complexity of this heuristic is $O(|N|\log^2|N|)$.

Table 3-1 Heuristic_LR_1 Algorithm

1.	Set each node i as inactive and assign it with weight $\max(0, \frac{\hat{a}_i(b_i) + \|N\|\mu_i^2}{(d_i)^2 + d_i/a_i})$. Sort all nodes by their weights in ascending order.
2.	Take source s as the root of the attack tree.
3.	Activate the first half of the inactive nodes.
4.	Use Prim's algorithm to construct the minimum cost sub-spanning tree for the activated nodes rooted at s.
5.	Examine each activated and uncompromised node. If its path cost is affordable for the attacker, apply sufficient attack power to compromise the node and all other uncompromised nodes on its path; then add all the nodes to the attack tree.
6.	Repeat Steps 3 to 5 until the attacker has insufficient attack resources to compromise any node.

3.3. Second-Stage Relaxation

Lagrangean relaxation problem

$$Z_{\mathrm{D}}(\nu_1, \nu_2, \nu_3) = \min_{y_i} \quad \textbf{(LR2)}$$

$$-\sum_{i\in N} d_i y_i + \sum_{i\in N} \nu_i^1 [\sum_{w\in W} \sum_{p\in P_w} x_p \delta_{pi} - (|N|-1) y_i] + \sum_{i\in N} \nu_i^2 [\sum_{p\in P_{(s,i)}} x_p - y_i] + \nu^3 [\sum_{i\in N} a_i - A],$$

subject to:
(IP 1.3) ~ (IP 1.6), (IP 1.8), and

$$u_i^1 \geq 0 \qquad \forall\, i \in N. \qquad \text{(LR 2.1)}$$

We decompose (LR 2) into three independent and easily solvable optimization subproblems with respect to decision variables x_p, and $[y_i, a_i]$, and solve the respective subproblems optimally.

Getting Primal Feasible Solutions

To improve the solution quality of (IP 1), we design and implement a heuristic while solving (LR 2). In this heuristic, each solution to (LR 2) is adjusted to a feasible solution to (IP 1). The basic concept of the heuristic, denoted as Heuristic_LR_2, is described below.

The subproblem related to variable x_p states that if the value of x_p is 1, an attack path is constructed, and all nodes on the path are targeted. By taking the union of constructed attack paths, we can form an attack tree. Then the attack tree can be adjusted to a feasible solution to (IP 1).

The time complexity of the first case is $O(|N|\log|N|)$, and that of the second case is $O(|N|^2)$.

Table 3-3 Heuristic_LR_2 Algorithm

1.	Assign each node i with weight $\max(0, \frac{\hat{a}_i(b_i) + \|N\|\mu_i^2}{(d_i)^2 + d_i/a_i})$.
2.	Examine all attack paths, i.e., paths whose value of x_p is 1, and add all nodes on the paths to the attack tree.
3.	Calculate the total cost of the resulting attack tree.
4.	If the total cost of the attack tree does not exceed the total attack budget:
4.1	Use Prim's algorithm to construct the minimum

```
        cost spanning tree on the basis of the current
        attack tree.
4.2     Find the uncompromised node with the smallest
        weight. Apply sufficient attack power to
        compromise the node if the attacker can
        construct an attack path to it. Then add the
        attack path to the attack tree.
4.3     Repeat Step 4.2 until the attacker has
        insufficient resources to compromise any node.
5.      If the total cost of the attack tree exceeds the
        total attack budget:
5.1     Find the leaf node of the attack tree with the
        largest weight. Remove it from the attack tree
        and withdraw the attack resource applied to it
        before.
5.2     Repeat Step 5.1 until the attack cost of the
        whole attack tree is affordable.
```

4. Computational Experiments

4.1. Computational Experiments with the AS Model

To demonstrate the effectiveness of the proposed heuristics, we implement the following algorithms for comparison purposes. The weight of each node is set to $\frac{\hat{a}_i(b_i)}{(d_i)^2}$ in the algorithms.

Simple Algorithm 1

The concept is derived from the heuristic of first-stage Lagrangean relaxation.

Simple Algorithm 2

The concept is based on the idea that nodes with smaller weights are more likely to be attacked. Here, we adopt Prim's algorithm to predetermine the path from *s* to each node.

Simple Algorithm 3

In order to compare the attack performance under conditions of complete and incomplete information, here we focus on the scenario where the attacker is only aware of the existence of uncompromised nodes through their compromised neighbors. The algorithm is based on the greedy method, and the total computational time of this heuristic is $O(|N|\log|N|)$.

4.2. Experiment Environment

The proposed algorithms for the AS model are coded in Visual C++ and run on a PC with an INTEL^{TM} Pentium 4.3GHz CPU. The parameters used in the experiments are detailed below.

Table 4-4 Experiment parameter settings for the AS model

Parameters	Value
Testing Topology	Grid (square), Random, Scale-free [4]
Number of Nodes $\|N\|$	100, 400, 900
Total Defense Budget	Equal to Number of Nodes
Total Attack Budget A	Equal to Total Defense Budget
Damage Distribution	Random distribution (D_1), Degree-based distribution (D_2), Uniform distribution (D_3)
Budget Allocation Strategy	Uniform allocation (B_1), Degree-based allocation (B_2), Damage-based allocation (B_3)
Defense Capability $\hat{a}_i(b_i)$	$\hat{a}_i(b_i) = 2b_i + \varepsilon$, b_i is the budget allocated to node i, $\forall\, i \in N$

4.3. Experiment Results

To compare attack behavior under different scenarios, we use the network susceptibility metric to evaluate the degree to which the attacker's objective is achieved. The greater the susceptibility, the more successful the attack. The LR value means the susceptibility calculated by the optimal feasible solution derived by the Lagrangean relaxation process. The LB value is a lower bound on LR, obtained from the smaller one of (LR 1) and (LR 2); and SA_1, SA_2, and SA_3 are the susceptibilities derived by simple algorithms 1, 2, and 3 respectively.

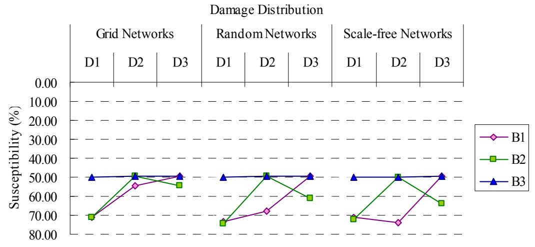

Figure 4-1 Susceptibility of medium-sized networks under different scenarios (|N| = 100)

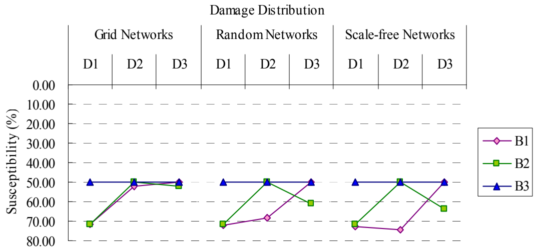

Figure 4-2 Susceptibility of large networks under different scenarios (|N| = 400)

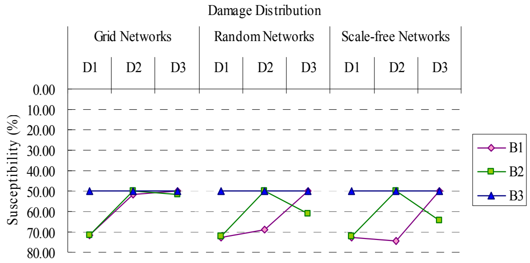

Figure 4-3 Susceptibility of extra-large networks under different scenarios (|*N*| = 900)

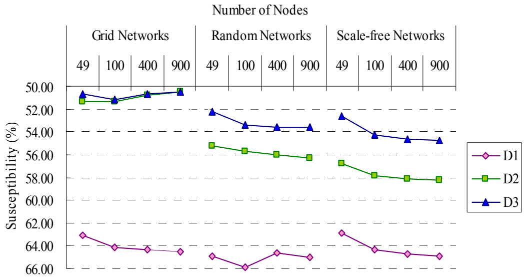

Figure 4-4 Susceptibility of different network sizes and damage distribution

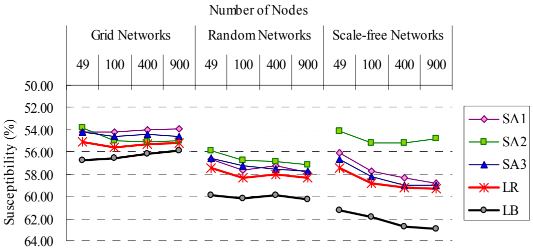

Figure 4-5 Susceptibility of different network sizes and topologies

4.4. Discussion of Results

Figures 4-1 to 4-4 show the susceptibility of the targeted network under different topology types, numbers of nodes, and damage distribution patterns. From these figures, we observe:

- Networks with budget allocation strategy B3 are the most robust and therefore the most difficult for an attacker to compromise. This finding is consistent with the common idea that defense resources should be allocated according to the importance of each node.
- For grid networks, the network susceptibility of the B1 and B2 strategies is close, and the gap between them decreases with the growth of the networks. This is because the degree of most nodes in a grid network is four, and the average degree approaches four when the network size increases.
- In random and scale-free networks, the average degree and the actual degree of each node diverge because of their randomness and power-law degree distribution characteristic [4] respectively. Thus, the B1 strategy, which treats each node equally, fails to reflect the discrepancy between the nodes, and results in high network susceptibility.
- Networks under the D3 scenario have the lowest susceptibility of the three damage distribution patterns. This result indicates that a network is more robust if "all nodes are created equal".

Figure 4-5 compares the solution quality of the proposed Lagrangean relaxation-based algorithm with simple algorithms 1, 2, and 3, and demonstrates the gap between LRs and LBs. From the figure, we observe:

- Our proposed heuristic outperforms the three simple algorithms in all cases. Our attack strategy causes the highest network susceptibility. This indicates that the proposed Lagrangean relaxation-based algorithm is not only capable of solving the AS model, it is also applicable to various types of network topology. The gaps between LRs and LBs are small, which shows that the proposed approach can derive a near-optimal solution to the AS model.
- Simple algorithm 2 performs very well in grid networks, but fails in scale-free networks. This strategy is only useful when there are multiple paths between the source and the target, as the attacker can make a detour when encountering nodes with a high defense capability. However, in the case of scale-free networks, the existence of "hubs", i.e., highly connected nodes [5], reduces the efficiency of simple algorithm 2, leading the rapid consumption of the attack budget.
- Simple algorithm 3 performs reasonably well in all types of network, especially scale-free networks. Due to this algorithm's local-information-awareness property, its solution quality is theoretically worse than that of the other algorithms. However, it turns out to be the opposite. One possible reason is that when an attacker has too much information, he may not be able to fully utilize it to develop a perfect attack strategy. On the other hand, the "six degrees of separation" property of scale-free networks allows an attacker to collect complete information about the targeted network once he has compromised several hub nodes.
- Generally, scale-free networks are more susceptible to attack than the other two topologies; grid networks are the least susceptible. Our finding that scale-free networks are more vulnerable to malicious attacks is consistent with previous research [5]. In contrast, the regular structure of a grid network makes it difficult for an attacker to compromise valuable nodes arbitrarily.

5. Conclusion and Future Work

5.1. Conclusion

The ubiquitous nature of the Internet has made it a magnet for cyber-crimes, which render the concept of "completely secure systems and networks" obsolete. Information theft is one of the most damaging cyber-crimes, yet it is easily missed because its attack behavior does not alert victims. Thus, in this paper we have considered the attack scenario in terms of information theft, where an attacker attempts to steal information from a targeted network and maximize his profit.

The key contribution of this work is that we successfully model the "silent" attack behavior into a well-formulated mathematical model, which is then solved by the proposed heuristic. This is a great stride in the topic of network attacks, since previous research seldom modeled real-world attack behavior in this way. Using mathematical forms, we can induce generic results and apply them to similar real-world scenarios that were only addressed by individual case studies in the past.

The novel network susceptibility metric is another contribution of this paper. The metric reflects the amount of profit gained by an attacker. This enables both the attacker and the defender to gauge the susceptibility of the targeted network and adjust their strategies accordingly. We have also studied several different network topologies and observed their susceptibility to information theft under different defense resource allocation strategies. The experiment results show that grid networks are the least susceptible to such theft, while scale-free network are the most susceptible.

5.2. Future Work

In this research, we adopt a linear defense capability function in the computational experiments. However, according to the "Law of Diminishing Marginal Utility", the marginal benefit, i.e., the additional defense capability derived from an additional unit of defense budget, declines as the defense budget increases. Thus, concave functions, e.g., log functions, may describe the real situation more accurately.

The current research only considers the best attack strategy under given defense strategies, but it would be more comprehensive if both strategies were considered simultaneously. Thus, the issue could be viewed as an offense-defense game and modeled as a two-level mathematical optimization problem, where the objective of the attacker is to maximize the total damage incurred by compromising nodes in a network, while the defender tries to minimize the total damage.

References

[1] A. Stewart, "On Risk: Perception and Direction," **Computers and Security**, Volume 23, pp. 362-370, May 2004.

[2] Y.-S. Lin, P.-H. Tsang, C.-H. Chen, C.-L. Tseng, and Y.-L. Lin, "Evaluation of Network Robustness for Given Defense Resource Allocation Strategies", **Proceedings of the 1st International Conference on Availability, Reliability and Security (ARES'06)**, pp. 182-189, April 2006.

[3] M.L. Fisher, "The Lagrangean Relaxation Method for Solving Integer Programming Problems," **Management Science**, Volume 27, Number 1, pp. 1-18, January 1981.

[4] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, Volume 286, pp. 509-512, October 1999.

[5] R. Albert, H. Jeong, and A.-L. Barabási, "Error and Attack Tolerance of Complex Networks," **Nature**, Volume 406, pp. 378-382, July 2000.