Cloud Computing Infrastructure Robustness: A Game Theory Approach

Nageswara S. V. Rao, Stephen W. Poole
Oak Ridge National Laboratory
Fei He, Jun Zhuang
State University of New York at Buffalo
Chris Y. T. Ma
Advanced Digital Sciences Center
David K. Y. Yau
Purdue University

***Abstract*—A cloud computing infrastructure typically consists of a number of sites that house servers and are connected to the Internet. Its operation critically depends both on cyber components, including servers and routers, and physical components, including fiber and power routes. Both types of components are subject to attacks of different kinds and frequencies, which must be accounted for the initial provisioning and subsequent operation of the infrastructure. The cyber and physical components may be individually attacked and defended, and the infrastructure is required to provide an aggregate computational capacity C. We present a game-theoretic approach for the provisioning and operation of the infrastructure under uniform cost models. We first show that the Nash Equilibrium under different formulations to be computable in polynomial time, and derive provisioning choices to ensure the capacity C with probability P_S. Then, we derive conditions for reinforcing the infrastructure, and show that higher robustness levels are achieved by limiting the disclosure of information about the infrastructure.**

I. Introduction

A cloud computing infrastructure typically consists of collections of computing servers deployed at multiple sites distributed over the Internet. At a basic level of abstraction, the infrastructure is required to provide a specified computing capacity aggregated among the currently available servers. A computational task may be executed on these servers, typically at locations unknown to users. The total computing power of all servers that are up and connected to the Internet is the *available capacity*, and is a primary performance measure. This infrastructure critically depends on the continued functioning of the cyber components, including computers and routers, as well as physical components, including fiber routes, cooling and power systems. These components may be degraded by deliberate cyber attacks on servers and routers as well as physical attacks on fiber and power routes. While cyber attacks on computing systems and networks seem to get more public media attention, in many occasions the infrastructure degradations have been due to physical factors such as fiber back hoe incidents. Indeed, this infrastructure can be compromised by attacking the physical components such as Heating, Ventilation and Air Conditioning (HVAC) systems, power-supply lines and physical fiber connections. The complexity of such physical attacks could be quite varied: attacks on HVAC systems require physical proximity access to the sites, whereas those on fiber and power routes can be anywhere along the stretches of unprotected areas that they run through. A physical disruption of the fiber or a cyber disruption of the gateway router of a site makes all servers unavailable, and a cyber attack on all servers of a site will have the same effect on the available capacity.

An infrastructure provider has to account for both cyber and physical attacks to ensure that the required capacity is available, both in the initial provisioning and also during the subsequent operation [4]. To account for the changing profiles of attacks and degradations during operation, the provider may reinforce the components: (i) cyber parts by replicating the servers and deploying fail-over gateway routers, and (ii) physical parts by using redundant, diverse fiber and power connections, and redundant HVAC systems. We present a game-theoretic formulation of the initial provisioning and subsequent operations of the infrastructure; the former deals with determining the number of servers to be deployed at different sites, and the latter deals with reinforcing selected physical and cyber components, both to ensure the required capacity. We consider the following conditions:

(a) knowledge about the number and physical locations of sites is available to attacker, primarily from the information provided to users;
(b) costs incurred by the provider and attacker are private information, and not available to the other;
(c) strategies used by the provider, namely the choice of servers to deploy at sites or components to reinforce, and by the attacker in choosing which parts to attack, are not revealed to the other.

Both provider and attacker consider that the other utilizes a probabilistic strategy. We consider discrete models consisting of servers, routers, sites and connections, which are much simpler than models used for critical infrastructures such as power distribution, transportation and agriculture [2, 7].

We consider that the utility functions of the attacker and provider are sums of cost and system terms, which they attempt to minimize [5, 8]. The Nash Equilibrium (NE) represents the attack and defense actions that minimize the utility of attacker and provider based on their

available information, respectively, from which neither has a motivation to unilaterally deviate [1, 3]. If costs depend only on the number of components, we show that NE can be computed with polynomial complexity in the number of components, from which the solutions to the initial provisioning and subsequent operations can be derived. We first derive sufficient conditions for selecting the number of servers at sites to ensure the capacity C with probability P_S, based on first order statistics of the attacker. We then derive NE conditions for reinforcing the infrastructure based on estimates of attack probabilities. The performance of infrastructure at NE depends on further details of reinforcement and attack strategies. We derive the expected capacity under statistical independence conditions. We show that the provider can hide and exploit the information about the distribution of servers across the sites to improve the expected capacity against the attacker.

In Section II, we describe a simplified cloud computing infrastructure model and solutions to infrastructure provisioning problem. We consider the reinforcement problem in Section III, and derive expected capacity estimates in Section IV.

II. Infrastructure Provisioning

We consider n_s sites, each connected to the Internet via fiber routes through a gateway router. Site i houses n_{s_i} servers each with a unit computing capacity. An attacker can launch cyber attacks on servers and gateway routers over the Internet, and physical attacks on fiber connections and physical plants. A physical attack on the fiber or physical plant, or a cyber attack on a gateway router will have essentially the same effect, namely, rendering all servers at the site unavailable. To simplify the discussion we collectively refer to these attacks as *site* or *physical* attacks. Let n_c and S_A, $|S_A| = n_s$ denote the randomly chosen number of cyber (server) attacks and set of sites unavailable due to physical or router attacks, respectively. Then the residual capacity is given by $\sum_{s_i \notin S_A} n_{s_i} - n_c$, which is the sum of capacity of all sites still connected to Internet minus number of servers compromised by cyber attacks. For the provider, the infrastructure provisioning problem is to determine n_{s_i}'s at different site to ensure a minimum capacity C with probability P_S. The probability that the capacity is at least C is given by

$$\begin{aligned}
&\mathbf{P}\left\{\sum_{s_i \notin S_A} n_{s_i} - n_c \geq C\right\} \\
&= 1 - \mathbf{P}\left\{n_c + C + \sum_{s_i \in S_A} n_{s_i} > \sum_{i=1}^{n_s} n_{s_i}\right\} \\
&\geq 1 - \frac{1}{n_t}\left(C + E[n_c] + E\left[\sum_{s_i \in S_A} n_{s_i}\right]\right),
\end{aligned}$$

where $n_t = \sum_{i=1}^{n_s} n_{s_i}$, and we have utilized the Markov's inequality, namely, $\mathbf{P}\{x > \epsilon\} < \frac{E[x]}{\epsilon}$, for positive x. Notice the Markov's inequality provides loose but useful bound for the initial provisioning. Then $E\left[\sum_{s_i \in S_A} n_{s_i}\right] = \bar{n}_s$ is the expected number of servers unavailable as a result of physical or router attacks. Let p_i denotes the probability of an attack on site i, then we have the alternative expression $\bar{n}_s = \sum_{i=1}^{n_s} p_i n_{s_i}$. Then, the capacity C can be assured with probability P_S by utilizing n_{s_i}'s values satisfying the equation $P_S = 1 - \frac{1}{n_t}\left(C + \bar{n}_c + \bar{n}_s\right)$, which equals to,

$$(1 - P_S)\sum_{i=1}^{n_s} n_{s_i} = C + \bar{n}_s + \bar{n}_c. \tag{2.1}$$

This solution depends on the expected number of servers attacked, and the expected size of site subject to attack. In the special case of all sites having the same number of servers $n_{s_i} = n_s$ we have a simpler formula $n_s = (C + \bar{n}_c)/[(1-P_S)n_s - 1]$ [6]. Since no conditions are needed for the Markov's bound, the resultant estimates could be quite high. Furthermore, high values of P_S could lead to impractically high cost of the infrastructure based on Eq (2.1) alone, which may not be met in certain cases.

Let $N_S = [n_{s_i}]$ denote the vector of all n_{s_i}'s. The cost of the infrastructure can be incorporated by minimizing the provider's expected utility function

$$\bar{U}_P(\bar{n}_c, N_S) = C_{P,S}\left(n_t\right) + C_{P,C}(\bar{n}_c) + \frac{C + \bar{n}_s + \bar{n}_c}{\sum_{i=1}^{n_s} n_{s_i}},$$

where the first and second terms on right hand side correspond to the costs of servers and sites, respectively, and third term is $1 - P_S$ given by Eq (2.1). The P_S attained by the minimization of this utility function by $N_S^* = [n_{s_i}^*]$ is

$$P_S = 1 - \frac{(C + \bar{n}_s + \bar{n}_c)}{n_t^*}, \tag{2.2}$$

where $n_t^* = \sum_{i=1}^{n_s} n_{s_i}^*$. This probability improves with total number of servers deployed n_t^*, and linearly degrades with the expected size of cyber attack $\bar{n}_c$ and expected number of servers at an attacked site.

Let $P_A = [p_i]$ denote the vector consisting of p_i's. We consider the expected utility function of the attacker given by sum as follows:

$$\begin{aligned}
\bar{U}_A(\bar{n}_c, P_A) &= C_{A,C}(\bar{n}_c) + C_{A,S}\left(\bar{n}_s\right) \\
&\quad + \mathbf{P}\left\{\sum_{s_i \notin S_A} n_{s_i} - n_c \geq C\right\} \\
&\leq C_{A,C}(\bar{n}_c) + C_{A,S}\left(\bar{n}_s\right) + \frac{1}{C^2}(\bar{n}_s - \bar{n}_c)^2,
\end{aligned}$$

where first and second terms represent the costs of attacking servers and site, respectively, under statistical independence, and third term corresponds to P_S. We utilized the Markov's inequality, namely, $\mathbf{P}\{X > \epsilon\} \leq E[X^2]/\epsilon^2$. The NE for this attacker is given by:

$$\frac{\partial C_{A,C}}{\partial \bar{n}_c} = \frac{2}{C^2}(\bar{n}_s - \bar{n}_c),$$

$$\frac{\partial C_{A,S}}{\partial \bar{n}_s} = -\frac{2}{C^2}(\bar{n}_s - \bar{n}_c).$$

The latter leads to a system of equations, which can be solved for P_A from which $\bar{n}_a$, $a = c, s$, can be estimated. If the provider has the knowledge of these values, they can be utilized in Eq (2.1)-(2.2).

III. Reinforcement Strategies

We consider that infrastructure has been operational and the provider now is required to reinforce the cyber and physical parts to withstand the degradations, which may have changed since initial provisioning. Let $x_a \geq 0$, $a = c, s$, be the number of components reinforced by the provider, and $y_a \geq 0$ be the number of the components attacked, where indexes $a = c$ and $a = s$ refer to server and site, respectively. The operational capacity of the infrastructure depends on the values of x_a and y_a, and the infrastructure requires $x_a - y_a \geq k_a$, $k_a \geq 0$ to be considered *operational*. The values for k_a's can be specified based on how strict the capacity requirement is. For example, $k_c = C$ and k_s^1 given by the $\min_k \sum_{i=1}^{k} n_{(s_i)} \geq C$ such that $n_{(s_1)}, \leq n_{(s_2)} \leq \cdots \leq n_{(s_{n_s})}$, will ensure that system has the capacity C when operational, and a smaller k_s may still provide capacity C. On the other hand, $k_c = C$ and k_s^2 given by $\min_k \sum_{i=1}^{k} n_{(s_{n_s - i})} \geq C$ will not guarantee the capacity C even if the infrastructure is operational. To simplify the presentation, we consider the more general formulation in terms of k_a's.

Our model is based on using $(n_c n_s + 1) \times (n_c n_s + 1)$ gain matrices, where rows represent the attacker choices and columns represent provider's options such that (i, j)th entry is interpreted as follows:

(a) For the attacker, bottom row $i = n_c n_s + 1$ represents attacking neither servers or sites, and $i \in [1, n_c n_s]$ represents attacking $i_s = (i-1) \div n_c + 1$ sites and $i_c = (i-1) \mod n_c + 1$ servers.

(b) For the provider, right most column $j = n_c n_s + 1$ represents defending neither servers nor sites, and $j \in [1, n_c n_s]$ represents defending $j_s = (j-1) \div n_c + 1$ sites and $j_c = (j-1) \mod n_c + 1$ servers.

We also denote (i, j)th entry $a_{i,j}$ in a more explicit form as $a_{i_c:i_s, j_c:j_s}$. The state vectors are given by

$$P_A = [p_{1:1}\ p_{1:2}\ \ldots\ p_{1:n_s} \ldots p_{n_c:1} \ldots p_{n_c:n_s}\ p\prime],$$
$$Q_D = [q_{1:1}\ q_{1:2}\ \ldots\ q_{1:n_s} \ldots q_{n_c:1} \ldots q_{n_c:n_s}\ q\prime],$$

where $p\prime = 1 - \sum_{i_c=1}^{n_c} \sum_{i_s=1}^{n_s} p_{i_c:i_s}$ and $q\prime = 1 - \sum_{j_c=1}^{n_c} \sum_{j_s=1}^{n_s} q_{j_c:j_s}$. For simplicity of notation, we also use the alternative notation $P_A = [p_1, p_2, \ldots, p_{n_c n_s + 1}]$ and $Q_D = [q_1, q_2, \ldots, q_{n_c n_s + 1}]$, where the index j represents j_c and j_s; q_j is also denoted by $q_{j_c:j_s}$. Let $c_{i,j}$, $d_{i,j}$ and $s_{i,j}$ denote the (i, j)th entry of cost the matrices of the attacker C^A, provider C^D, and the system matrix R^A, respectively. The bottom row of C^A consists of 0's denoting the cost of no attack with probability $p\prime$, that is $c_{n_c n_s + 1, j} = 0$ for $j = 1, 2, \ldots, n_c n_s + 1$. And the right-most column of C^D consists of 0's denoting no reinforcement, that is $d_{i, n_c n_s + 1} = 0$ for $i = 1, 2, \ldots, n_c n_s + 1$.

Under the condition of uniform server and site costs, the cost matrices may be specified as follows: (a) for provider $d_{i,j} = d_{i,j_c:j_s} = j_c d_{dc} + j_s d_{ds}$, where d_{dc} and d_{ds} are costs of reinforcing server and site, respectively, and (b) for attacker $c_{i,j} = i_c c_{ac} + i_s^{\alpha} c_{as}$ where c_{ac} and c_{as} are the costs of attacking server and site, respectively; α could be higher than 1 indicating the higher cost of coordinating multiple site attacks at geographically separated locations.

We consider that the utility functions of the attacker and provider consists of sum of cost and system matrices, given by $G^A = C^A + R^A$ and $G^D = C^D - R^A$, respectively. The cost term for the attacker is $P_A C^A Q_D^T$ where Q_D^T represents the attacker's estimate of provider's probabilities of reinforcement. The system term utilized by the attacker is $P_A R^A Q_D^T$. At Nash Equilibrium, attacker computes P_A^* that minimizes $P_A G^A Q_D^T$, and provider computes Q_D^* that minimizes $P_A G^D Q_D^T$. For the attacker, the partial derivative with respect to p_i is

$$\frac{\partial P_A G^A Q_D^T}{\partial p_i} = q'(s_{i, n_c n_s + 1} - s_{n_c n_s + 1, n_c n_s + 1}) + \sum_{j_c=1}^{n_c} \sum_{j_s=1}^{n_s} q_{j_c:j_s} \left(c_{i,j_c:j_s} + s_{i,j_c:j_s} - s_{n_c n_s + 1, j_c:j_s}\right),$$

Then NE is determined by computing all above partial derivatives that are negative, and assigning probability 1 to the one that minimizes $P_A G^A Q_D^T$. Since each of these terms is based on "fixed" elements of the gain matrices and no limits are imposed on the cost, the corresponding probability can be increased to 1. However, if the elements depend on the probabilities, this approach does not result in the minimization of utility function. If all partial derivative are non-negative, then attacker will not attack, i.e., $p\prime = 1$, and the system survives. The computational complexity of this step is $O\left(n_c^2 n_s^2\right)$. This computation requires q_j's, which are attacker's estimates of the probabilities of components being reinforced. Such information can be based on public information and best practices.

For the provider we have cost term $P_A C^D Q_D^T$ and combining with the system term $P_A R^D Q_D^T = -P_A R^A Q_D^T$, we have

$$\frac{\partial P_A G^D Q_D^T}{\partial q_j} = \sum_{i=1}^{n_c} p_i (d_{i,j} - s_{i,j} + s_{i, n_c n_s + 1}) + p'(d_{n_c n_s + 1, j} - s_{n_c n_s + 1, j} + s_{n_c n_s + 1, n_c n_s + 1}).$$

Here P_A represents the provider's estimate of the attacker's probabilities. Then, we compute all the resultant terms that are negative, and pick the one that gives the lowest cost for $P_A G^D Q_D^T$. If no negative partial derivatives exist, no components will be reinforced, i.e. $q\prime = 1$, and the system may not be operational after the attack. Thus at NE, the system's operational status is deterministic as follows:

$$\text{system state} = \begin{cases} \text{operational} & \text{if } [(x_c \geq k_c) \wedge (x_s \geq k_s)] \\ & \vee[(y_c < n_c - k_c) \wedge (y_s < n_s - k_s)] \\ \text{not} & \text{else if} \\ & [(x_c < k_c) \wedge (y_c > n_c + x_c - k_c)] \\ & \vee[(x_s < k_s) \wedge (y_s > n_s + x_s - k_s)] \\ \text{either} & \text{else} \end{cases}$$

The system status in the third case depends on which components are attacked and reinforced. For $a = c, s$, there are less than k_a components reinforced, and no more than $n_a - y_a$ components not attacked, since $x_a < k_a$ and $n_a - k_a \geq y_a \leq n_a + x_a - k_a$ Thus, there is set $S_{n_a - y_a}$ with at least $n_a - y_a \leq k_a$ components not attacked, and there is a set S_{x_a} with $x_a < k_a$ components that are reinforced. The system will remain operational if and only if there are k_a, for $a = c, s$, components each of which is either not attacked or has been reinforced, that is $|S_{n_a - y_a} \cup S_{x_a}| \geq k_a$, for $a = c, s$. If the system is operational, its performance level in terms of the residual reinforcements is determined by x_a, y_a, n_a and k_a, for $a = c, s$, and also the precise strategies used by the attacker and provider as illustrated in the next section.

IV. Operational Capacity Estimation

Let $i_c = y_c$ and $i_s = y_s$ denote the number of server and sites attacked, respectively, and $x_c = j_c = (j-1) \bmod n_c + 1$ and $x_s = j_s = (j-1) \div n_c + 1$ denote the number of cyber and physical components reinforced by the provider, respectively. We consider two ways of computing element $s_{i,j}$ of the system matrix R^A, which in turn determines the values of x_a and y_a. We first consider that system terms given by

$$s^I_{i,j} = \begin{cases} 2S & \text{if } [(y_c = 0) \wedge (y_s = 0)] \\ -2S & \text{else if} \\ & [(x_c < k_c) \\ & \quad \wedge (y_c > n_c + x_c - k_c)] \\ & \vee[(x_s < k_s) \\ & \quad \wedge (y_s > n_s + x_s - k_s)] \\ S\left[1 + \frac{(x_c - k_c)(x_s - k_s)}{(x_c - k_c + y_c)(x_s - k_s + y_s)}\right] & \text{else} \end{cases}$$

In the first case, there is no attack hence the system survives at the reinforced level. In the next case, the system will not survive since the required number of cyber and physical components are not available. In the last case, the system operates with a degraded capacity, and the residual capacity is proportional to $\frac{1}{y_a}$, $y_a = 1, 2, \ldots, n_a$, $a = c, s$, reflecting the probability of attack under statistical independence condition.

We consider another way to specify the system terms, where residual capacity is proportional to $-y_a$: the last case above is given by

$$s^{II}_{i,j} = S\left[1 + \frac{(x_c - k_c - y_c)(x_s - k_s - y_s)}{(x_c - k_c)(x_s - k_s)}\right],$$

and the other cases are identical to $s^I_{i,j}$.

NE conditions specify only the values of x_a and y_a, $a = c, s$, and the actual choice of which components to attack and reinforce depends on the strategies used by the provider and attacker. We now estimate the expected residual capacity when the attacker and provider pick the components to attack and reinforce, respectively, independently using uniform distribution, and the attack and reinforcement probabilities are statistically independent. A cyber or physical component will survive if it is not attacked or has been reinforced when attack occurs; for $a = c, s$, these probabilities are given by $(1 - 1/n_a)^{y_a}$ and $[1 - (1 - 1/n_a)^{x_a}][1 - (1 - 1/n_a)^{y_a}]$, respectively, where $n_c = \sum_{i=1}^{n_s} n_{s_i}$. The probability that a component will survive an attack is given by, for $a = c, s$,

$$1 + (1 - 1/n_a)^{x_a}(1 - 1/n_a)^{y_a} - (1 - 1/n_a)^{x_a}.$$

The attacker will attack y_c servers distributed among the sites. On the other hand, each physical attack is on a single site, which will disconnect all the servers at the site.

Consider that provider will adopt a strategy of assigning higher reinforcement probabilities to sites with higher number of servers, namely probability $\frac{n_{s_i}}{\sum_{j=1}^{n_s} n_{s_j}}$ for a site with n_{s_i} servers. On the other hand, consider that attacker will adopt a uniform strategy being unaware of the number of servers at different sites. Then, the probability that any node will be selected for reinforcement and attack are given by $1 - \left(1 - \frac{n_{s_i}}{\sum_{j=1}^{n_s} n_{s_j}}\right)^{x_a}$ and $1 - (1 - 1/n_a)^{y_a}$, respectively. Then the expected operational capacity $\hat{C}_L$ for this linear strategy is

$$\sum_{i=1}^{n_s}\left(n_{s_i}\left[1 - \sum_{a=c,s}\left(1 - \frac{n_{s_i}}{\sum_{j=1}^{n_s} n_{s_j}}\right)^{x_a}\left[1 - \left(1 - \frac{1}{n_a}\right)^{y_a}\right]\right]\right).$$

The expected capacity $\hat{C}_U$ for the uniform strategy is

$$\sum_{i=1}^{n_s}\left(n_{s_i}\left[1 - \sum_{a=c,s}\left(1 - \frac{1}{n_{s_i}}\right)^{x_a}\left[1 - \left(1 - \frac{1}{n_a}\right)^{y_a}\right]\right]\right).$$

The inequality $\left(\sum_{j=1}^{n_s} n_{s_j}\right)^2 \leq n_s \sum_{i=1}^{n_s} n_{s_i}^2$, is sufficient to ensure that $\hat{C}_L > \hat{C}_U$. Thus this provider's approach ensures a higher expected capacity compared to a uni-

Case	Parameters						Simulation Results			
	k_c	k_p	c_{ac}	c_{as}	d_{dc}	d_{ds}	attack	defense	survival	residual capacity
A.	5	3	1	1	1	1	30 (c), 5(s)	5(c), 3(s)	100% (both)	9.82 (prop), 8.35(uni)
B.	5	3	100	1	1	1	1 (c), 5(s)	5(c), 3(s)	100% (both)	22.78 (prop), 18.75(uni)
C.	5	3	100	1	1	100	1(c), 5(s)	1(c), 0(s)	0% (both)	0 (both)
A'.	5	3	1	1	1	1	30 (c), 5(s)	5(c), 3(s)	100% (both)	5.21 (uni)
B'.	5	3	100	1	1	1	1 (c), 5(s)	5(c), 3(s)	100% (both)	18.9(uni)

TABLE I
SIMULATION OF 30 SERVER CLOUD COMPUTING INFRASTRUCTURE; C AND P DENOTE CYBER AND PHYSICAL PARTS, AND PROP AND UNI DENOTE PROPORTIONAL AND UNIFORM STRATEGIES FOR THE DEFENDER.

form strategy. However, if the provider discloses n_{s_i}'s, then the attacker might adopt a less uniform strategy. Thus by not disclosing this information, the provider gains a definite advantage as the attacker is made to adopt a less informed strategy.

We simulated a cloud computing infrastructure with 30 servers distributed at 5 sites with various parameters using the system term $s^I_{i,j}$ (results are qualitatively quite similar under $s^{II}_{i,j}$). Given the set of parameters, $k_c, k_p, c_{ac}, c_{as}, d_{dc}, d_{ds}$, the equilibria of reinforcement and attack are obtained based on formula in Section III. Then, the survival probability and residual capacity are calculated using the above formulae. At NE, we compute the system status and the available capacity by simulating 1000 instances of the attacker and provider strategies; for the latter, we consider both uniform and proportional methods. The salient features of the simulations are summarized in Table I. We consider two server distributions of 15, 10, 3, 1 and 1, and all 6 servers across the sites. In cases A-C, we consider non-uniform server distribution. In case A, the attacker attacks both servers and sites since their costs are the same, and the defender defends both, and hence the system survives. In case B, the cyber attack cost is much higher leading a smaller number of server attacks, and consequently, the residual capacity is much higher. In both cases A and B, the proportional defense strategy yields a higher expected capacity. In case C, the cost of site defense is much higher and consequently the defender does not choose to defend the sites; as a result, the physical attacks bring the system down completely. These cases illustrate that dominant costs strongly influence not only the capacity of the infrastructure but also its very survival. Cases A' and B' are identical to A and B, respectively, except they use the uniform server distribution. The defender has no particular advantage over the attacker in this case, and as a result the capacity is lower compared to the corresponding cases A and B.

V. CONCLUSIONS

We presented a game-theoretic approach to the problem of provisioning and reinforcing cloud computing infrastructures, which provided insights into the qualitative effects of costs and strategies. We consider this work to be an initial step that can be extended in several ways. The basic results of this paper can be extended to account for incidental degradations and probabilities of successful attacks and reinforcements by suitably augmenting the state vectors [6]. Clearly, these models can be extended by considering servers of different capacities and sites with limitations on connectivity and physical space. The simple Internet connectivity model can be expanded to consider complex network topologies with LAN switches and core routers in the WAN. It would be interesting to study sequential game formulations of this problem, and cases where different levels of knowledge are available to the other party. More detailed simulations with system-specific details would be future interest.

ACKNOWLEDGMENTS

This work is funded by the Mathematics of Complex, Distributed, Interconnected Systems Program, Office of Advanced Computing Research, U.S. Department of Energy, and by the Extreme Scale Systems Center, sponsored by U. S. Department of Defense, and is performed at Oak Ridge National Laboratory managed by UT-Battelle, LLC for U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

REFERENCES

[1] T. Alpcan and T. Basar. *Network Security: A Decision and Game Theoretic Approach*. Cambridge University Press, 2011.

[2] G. Brown, M. Carlyle, J. Salmern, and K. Wood. Defending critical infrastructure. *Interfaces*, 36(6):532–544, 2006.

[3] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 2003.

[4] E. A. Lee. Cyber physical systems: Design challenges. In *International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC)*, 2008.

[5] C. Y. T. Ma, N. S. V. Rao, and D. K. Y. Yau. A game-theoretic study of attack and defense in cyber-physical systems. In *International Workshop on Cyber-Physical Networking Systems*, 2011.

[6] N. S. V. Rao, D. K. Y. Yau, and C. Y. T. Ma. On robustness of cyber-physical network infrastructure. In *Workshop on Design, Modeling and Evaluation of Cyber Physical Systems*. 2011.

[7] C. W. Ten, G. Manimaran, and C. C. Liu. Cybersecurity for critical infrastructures: attack and defense modeling. *IEEE Transactions on System, Man and Cybernectics: Part A: Systems and Humans*, 40(4):853–865, 2010.

[8] J. Zhuang and V. M. Bier. Balancing terrorism and natural disaters - Defensive strategy with enddogenous attacker effort. *Operations Research*, 55(5):976–991, 2007.