When Selfish Meets Evil: Byzantine Players in a Virus Inoculation Game

Thomas Moscibroda
Computer Engineering and Networks Laboratory
ETH Zurich
8092 Zurich, Switzerland
moscitho@tik.ee.ethz.ch

Stefan Schmid
Computer Engineering and Networks Laboratory
ETH Zurich
8092 Zurich, Switzerland
schmiste@tik.ee.ethz.ch

Roger Wattenhofer
Computer Engineering and Networks Laboratory
ETH Zurich
8092 Zurich, Switzerland
wattenhofer@tik.ee.ethz.ch

ABSTRACT

Over the last years, game theory has provided great insights into the behavior of distributed systems by modeling the players as utility-maximizing agents. In particular, it has been shown that selfishness causes many systems to perform in a globally suboptimal fashion. Such systems are said to have a large Price of Anarchy. In this paper, we extend this active field of research by allowing some players to be malicious or Byzantine rather than selfish. We ask: What is the impact of Byzantine players on the system's efficiency compared to purely selfish environments or compared to the social optimum? In particular, we introduce the Price of Malice which captures this efficiency degradation. As an example, we analyze the Price of Malice of a game which models the containment of the spread of viruses. In this game, each node can choose whether or not to install anti-virus software. Then, a virus starts from a random node and iteratively infects all neighboring nodes which are not inoculated. We establish various results about this game. For instance, we quantify how much the presence of Byzantine players can deteriorate or—in case of highly risk-averse selfish players—improve the social welfare of the distributed system.

Categories and Subject Descriptors

F.2 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity

General Terms

Theory, Economics

Keywords

Selfishness, Game Theory, Price of Anarchy, Price of Malice, Byzantine Nash Equilibria, Virus Propagation

1. INTRODUCTION

The introduction of micro economic models in computer science has led to great insights into the reality of today's distributed systems such as the Internet, which typically connect selfish, utility-optimizing agents or *players*. Over the last years, many aspects of distributed systems have been studied from a game-theoretic point of view. A particularly exciting question concerns the so-called *Price of Anarchy*: How much better would the social welfare be if the selfish players collaborated instead of seeking to maximize their own benefit?

However, selfishness is not the only challenge to the performance of distributed systems. Often, these systems have to cope with malicious *Byzantine adversaries* who seek—independently of their own cost—to degrade the utility of the entire system, to attack correctness of certain computations, or to cause endless changes which render the system instable. Aware of these threats, many researchers especially in the area of security and distributed computing have devised solutions to defend against such possible attacks.

In this paper, we aim at combining these two fruitful threads of research. In particular, we consider a system of selfish individuals whose only goal is to optimize their own benefit, and add malicious players who attack the system in order to deteriorate its overall performance. We ask: What is the impact of the Byzantine players on a selfish system's efficiency?

We believe that this question is of interest and actuality in many research fields. Examples in computer science include *Internet viruses* or *Denial of Service attacks* where some players aim at destructing systems which otherwise typically consist of utility-maximizing players. However, such phenomena might also arise in economic or sociological environments. For instance, one can imagine a set of companies competing on a market, selfishly seeking to maximize their individual gains. Among them, there might be one or two companies run by "terrorists" whose goal is to destabilize the economic system.

In order to capture these questions formally, we introduce the *Price of Malice* of selfish systems. The Price of Malice is a ratio that expresses how much the presence of malicious players deteriorates the social welfare of a system consisting of selfish players. More technically, the Price of Malice is the ratio between the social welfare or performance achieved by a selfish system containing a number of Byzantine players, and the social welfare achieved by an entirely selfish society.

It is interesting to compare the Price of Malice with the notion of the Price of Anarchy. The Price of Anarchy captures the degradation of a socially optimal performance of a system due to selfish behavior of its users or participants. That is, the Price of Anarchy relates the social welfare generated by players acting in an egoistic manner to an optimal solution obtained by perfectly collaborating participants. The Price of Malice's reference point, on the other hand, is not a socially optimal welfare, but the welfare achieved by an entirely selfish system.

PODC'06, July 22-26, 2006, Denver, Colorado, USA.

The Price of Anarchy and the Price of Malice are therefore two orthogonal measures that describe inherent properties of distributed, socio-economic systems. Specifically, a system may have a small Price of Anarchy, but a large Price of Malice, and vice versa. The fact that a system has a large *Price of Anarchy* indicates that it is necessary to design mechanisms (such as taxes or payment schemes) that forces players to collaborate more effectively. However, it is much more difficult to improve (or *repair*) systems having a large *Price of Malice*, since Byzantine players do not respond to any rules or (financial) incentives. Often, the only solution is to defend the system against malicious intruders, or at least to ensure that the number of malicious players in the system remains small.

By introducing a model that comprises the notions of *Byzantine Nash equilibria*, the *Byzantine Price of Anarchy*, and the *Price of Malice*, we are able to formally analyze what happens in selfish systems if one or more players' aim is to hinder the system from working or to bog down its performance as much as possible.

The Price of Malice crucially depends on the amount of information the selfish players have about the presence and behavior of the Byzantine players, and how they respond to this information. In other words, the utility function which finally defines the selfish players' reaction depends on how they *subjectively* perceive and judge the threat of Byzantine players. Hence, the utility of selfish players is computed using the *perceived expected cost* rather than the unknown actual cost. For example, it can be shown that in case of risk-averse players, the presence of Byzantine players may actually *improve* the social welfare compared to a situation where there are no Byzantine players at all. That is, there are situations where selfish players may tend to be more willing to collaborate if they face higher risks.

To the best of our knowledge, this is the first paper to study a model which allows for an *analytical quantification* of this so-called *Fear Factor*. Potentially, this in turn gives raise to several research questions in many areas including distributed systems, economics, politics or sociology.

Besides studying Byzantine players who aim at minimizing the performance of a system (Price of Malice), we also raise the question of *stability*. Particularly, we are interested in the question, how many Byzantine players suffice in order to prevent the system from stabilizing.

In this paper, we investigate a concrete example where selfish and Byzantine players interact. In this simple game, we consider a network of nodes, where each node can choose between paying for inoculation, or risking to get infected by a virus. After the nodes have made their choices, a virus starts at some random node and propagates iteratively to all neighboring nodes which are not inoculated. For this game, we give tight bounds for the Price of Malice for different types of selfish nodes (not aware of Byzantine nodes, aware and risk-averse, etc.).

The remainder of this paper is organized as follows. After reviewing related work in Section 2, we introduce the foundations of our Byzantine game theory in general and the models of the virus inoculation game in particular in Section 3. In Section 4, the impact of selfish and Byzantine behavior on the social welfare of the virus inoculation game is studied. Section 5 considers Byzantine attacks on the *stability* of a system, before Section 6 concludes the paper.

2. RELATED WORK

Security and robustness of distributed systems against Byzantine faults have been of prime importance and an active field of research for many years. Possibly the most well-known problem in this context has been that of reaching *consensus* among distributed parties. Possibility and impossibility results on the Byzantine consensus problem have been achieved in a variety of models and settings. Classic work in the synchronous and asynchronous case includes [4, 12, 21] and [9], respectively. In addition to the consensus problem, the distributed computing community has come up with results and solutions for a wide variety of other problems with Byzantine faults. Examples are clock synchronization [23], broadcast [10, 22], or quorum systems [13]. All of the above works assume that non-Byzantine players (or processes) are benevolent and attempt to reach a common goal. Finally, Byzantine behavior is subject to intensive research in cryptography. For instance, there is a large body of work in the area of secure multi-party computation [24].

In his STOC'01 talk [17], Papadimitriou has argued that the Internet has surpassed the von Neumann computer as the most complex computational artifact of our time. In particular, he pointed out that the Internet has a *socio-economic* complexity whose understanding requires techniques from mathematical economics and game theory [16]. Since then, game theoretic approaches have become increasingly popular to study selfish behavior on all layers of distributed systems. Specifically, researchers have been keen to study the inherent loss of efficiency in a system caused by the participant's selfishness in networks. Consequently, the Price of Anarchy [11, 20] and its complexity has been investigated in various system settings, including the Internet [7], wireless ad-hoc networks [5], or peer-to-peer systems [14]. Enforcing a truthful behavior or a reasonable efficiency in systems with a potentially high Price of Anarchy has been the goal of *algorithmic mechanism design* [8, 15].

In this paper, we strive for combining these two threads of research. In this respect, our research is related to the notions of *fault tolerant implementation* introduced by Eliaz [6] and of *BAR fault tolerance* introduced by Aiyer et al. [1]. In [6], implementation problems are investigated where there are k faulty players in the population, but neither their number nor their identity is known. A planner's objective then is to design an equilibrium where the non-faulty players act according to his rules. In [1], the authors describe an asynchronous state machine replication protocol which tolerates Byzantine, Altruistic, and Rational behavior. Interestingly, they find that the presence of Byzantine players can simplify the design of protocols if players are risk averse.

There exists other work on game theoretic systems in which not every participating agent acts in a rational or selfish way. In *Stackelberg theory* [19], for instance, the model consists of a set of selfish players, but a certain fraction of the entire population is *controlled by a global leader*. The leader's goal is to devise a strategy that induces an optimal or near optimal so-called Stackelberg equilibrium.

Virus propagation models have also been widely studied in literature. While traditional epidemiological models characterize infection in terms of birth rate and death rate of the virus [3], more recently models have been proposed for all kind of graphs, including Internet-like power-law graphs [18]. In particular, the percolation and game theoretic virus propagation model of this paper is based on [2]. The authors of [2] model the containment of the spread of viruses in general graphs. They characterize equilibria in selfish environments and also give an approximation algorithm for the centralized, non-selfish case.

3. MODEL

We present our model in two steps. First, we discuss the virus inoculation game derived from [2]. Subsequently, we introduce our framework of Byzantine game theory including the definition of the Price of Malice.

3.1 Virus Inoculation Game

Similarly to [2], we model the virus inoculation game as a scenario with n strategic players each of whom corresponds to a node

in an undirected grid $G[r, c]$ of r rows and c columns.[1] Henceforth, we will refer to the upper left corner of the grid as $G[0, 0]$, i.e., indices start with 0.

Each node i has two choices: either do nothing and risk infection by a virus, or inoculate itself by installing anti-virus software. For a node, installing the anti-virus software has the obvious advantage that it becomes immune against infection. On the other hand, the process of installing the software entails a cost in terms of money and/or time. Hence, a strategic player may or may not opt for inoculation depending on which choice maximizes his own utility.

The nodes' choices can be summarized by a strategy profile $\overrightarrow{a} \in \{0, 1\}^n$, where $a_i = 1$ signifies that node i installs the anti-virus software, and $a_i = 0$ that it does not install it. We call nodes i with $a_i = 1$ *secure*, and denote the set of secure nodes by $I_{\overrightarrow{a}}$. After the nodes have made their choices, the adversary picks some node uniformly at random as a starting point for infection. Infection then propagates on the network graph and infects all non-secure nodes that are in the same non-secure connected component as the starting point of infection. Technically, we associate an *attack graph* $G_{\overrightarrow{a}} = G \setminus I_{\overrightarrow{a}}$ with $\overrightarrow{a}$. It is essentially the network graph in which all secure nodes and their incident edges are removed.

In this paper, we consider the following costs: installing anti-virus software on a selfish node entails an inoculation cost of 1 at this node. If a selfish node does not inoculate and becomes infected, it suffers a loss equal to L. Therefore, the cost of a selfish node i can be summarized as follows:

$$cost_i(\overrightarrow{a}) = a_i + (1 - a_i) \cdot L \cdot \frac{k_i}{n}, \tag{1}$$

where k_i/n is the probability that node i is infected, conditioned on the event that it does not install the anti-virus software. Thereby, k_i is the size of the connected component containing i in $G_{\overrightarrow{a}}$. Finally, the *social cost* of a strategy profile $\overrightarrow{a}$ is the sum of all individual costs, i.e., $Cost(\overrightarrow{a}) = \sum_{j \in \mathcal{S}} cost_j(\overrightarrow{a})$, where S denotes the set of all selfish players. When the strategy profile $\overrightarrow{a}$ is clear from the context, we sometimes use abbreviations $cost_i$ and $Cost$ to denote individual cost and social cost, respectively.

3.2 Byzantine Game Theory

In order to understand the impact of malicious players on the selfish system, we extend the virus inoculation game with malicious Byzantine players. Formally, there are n nodes in the network. Of these n nodes, b are malicious *Byzantine nodes* that do not strive for minimizing their own costs. Instead, the goal of these Byzantine nodes is to deteriorate the overall system performance as much as possible, i.e., to maximize the resulting social cost of the solution. The remaining $s := n - b$ nodes are *selfish* and aim at maximizing their own utility. Instead, the goal of these Byzantine nodes is to deteriorate the overall system performance as much as possible, i.e., to maximize the resulting social cost of the solution. We denote the set of Byzantine and selfish players as B and S, respectively. It holds that $b := |B|$, $s := |S|$, and $n = s + b$.

While selfish nodes behave as discussed in Section 3.1, we assume that the Byzantine nodes pursue the following strategy: they claim to be inoculated (i.e., they proclaim their strategy to be $a_i = 1$), but actually they are not. In order to emphasize that Byzantine nodes are only seemingly secure, we denote the set of really inoculated and secure selfish nodes by $I^{self}_{\overrightarrow{a}}$. The attack graph resulting from strategy profile $\overrightarrow{a}$ is then $G_{\overrightarrow{a}} = G - I^{self}_{\overrightarrow{a}}$. This is the network graph without secure, selfish nodes, but including all Byzantine nodes. We can therefore define the individual cost incurred at a selfish node $i \in S$ as follows.

DEFINITION 3.1 (ACTUAL INDIVIDUAL COST). *The (actual) individual cost $cost_i(\overrightarrow{a})$ of a node $i \in S$ is defined as*

$$cost_i(\overrightarrow{a}) := a_i + (1 - a_i) \cdot L \cdot \frac{k_i}{n},$$

where k_i is the size of the connected component of node i in the attack graph $G_{\overrightarrow{a}}$.

Notice that in spite of its being equivalent to the corresponding definition in Section 3.1, we call this cost *actual individual cost*. This is to emphasize the fact that selfish players may not know about the existence of Byzantine players, and therefore, they are unable to compute their actual individual cost. Even if they are aware of the malicious players' existence, they might not know the Byzantine players' exact locations or strategies. In other words, with the addition of Byzantine players, selfish nodes no longer have a *perfect knowledge* about the network and its nodes' choices.

In case of imperfect information, a node might deal with its uncertainty in different ways. For example, a node might be risk averse and act in a conservative manner. These observations imply that before the location and strategies of Byzantine players are revealed (i.e., before the virus infection occurs), a selfish player i experiences a *perceived individual cost* $\widehat{cost}_i(\overrightarrow{a})$. This perceived cost can differ from the *actual individual cost* $cost_i(\overrightarrow{a})$ a node eventually has to pay.

DEFINITION 3.2 (PERCEIVED INDIVIDUAL COST). *Consider a selfish game with Byzantine players in which selfish players have imperfect knowledge about the existence, location, or the strategy of Byzantine players. In this case, the perceived individual cost $\widehat{cost}_i(\overrightarrow{a})$ of a selfish player i captures the cost expected by player i given his knowledge about the Byzantine players. This cost depends on the underlying model.*

The strategic decisions of selfish players can only be based on the *perceived cost* (not on their actual individual costs), as the actual individual cost can only be computed once the locations and strategies of Byzantine players are revealed. In this paper, we will study the following two basic models.

DEFINITION 3.3 (OBLIVIOUS). *In the* oblivious model*, selfish players are not aware of the existence of Byzantine players. That is, selfish players assume that all other players in the system are selfish as well.*

DEFINITION 3.4 (NON-OBLIVIOUS). *In the* non-oblivious model*, selfish players know about the existence of Byzantine players. Specifically, we assume that every selfish player knows b, the number of Byzantine players in the system, but he does not know about these players' exact locations or strategies. Moreover, we assume that selfish players are highly risk averse in the sense that they aim at minimizing their maximal individual cost. Let $\mathcal{D}$ be the set of possible distributions of Byzantine players among all players. A selfish player i experiences a perceived individual cost of*

$$\widehat{cost}_i(\overrightarrow{a}) := \max_{d \in \mathcal{D}} \{cost_i(\overrightarrow{a}, d)\},$$

where $cost_i(\overrightarrow{a}, d)$ denotes the actual costs of i if the Byzantine players are distributed according to $d \in \mathcal{D}$.

In the virus inoculation game, and in an oblivious model, the perceived cost is typically smaller than the actual cost: A node $i \in S$ does not take into consideration the Byzantine nodes which may

[1]Our results can be generalized to other highly regular, low-dimensional graphs such as the two-dimensional torus, i.e., a grid that wraps around at the boundaries.

increase the size of i's attack component. In the non-oblivious risk-averse model on the other hand, a node actually overestimates its expected actual cost by considering a worst-case scenario: A selfish player assumes that the Byzantine nodes are—from its individual point of view—distributed in a worst-case fashion among all players. Therefore, the perceived individual cost may be larger than the actual cost.

Since our goal is to understand the impact of terrorist behavior on a system of selfish players, the cost of Byzantine players is not included in the social cost. If it was, it would in general be easy for Byzantine players to arbitrarily deteriorate the social welfare of a system by simply increasing their own costs as much as possible. Moreover, as Byzantine players are malicious anyway, there is no particular reason why the overall system should care about these players' costs.

Formally, the total *social cost* $Cost(\overrightarrow{a})$ of a strategy is defined as the sum of the (actual) individual costs of all selfish players. Since each node in the same connected component of $G_{\overrightarrow{a}}$ has the same probability of infection, the l_i selfish nodes in the i-th attack component face a loss of $l_i \cdot (Lk_i/n)$ if the component is infected.

DEFINITION 3.5 (SOCIAL COST). *The social cost is given by the sum of the actual individual costs of selfish players*

$$Cost(\overrightarrow{a}) \;=\; \sum_{j \in \mathcal{S}} cost_j(\overrightarrow{a}) \;=\; \underbrace{|I^{self}_{\overrightarrow{a}}|}_{\text{inoculation cost}} \;+\; \underbrace{\frac{L}{n} \sum_{i=1}^{l} k_i l_i}_{\text{infection cost}},$$

where $k_1, k_2, \ldots, k_l$ are the sizes of the components in $G_{\overrightarrow{a}}$, and $l_1, l_2, \ldots, l_l$ are the sizes of the same components without counting the Byzantine nodes. We refer to the cost due to inoculation as the inoculation cost $Cost_{inoc}$*, and to the cost due to the virus infections as the* infection cost $Cost_{inf}$.

As customary in the game theory literature [16], the social cost of a setting where all nodes perfectly collaborate, i.e., where there are neither selfish nor Byzantine nodes, is called the *social optimum*.

DEFINITION 3.6 (OPTIMAL SOCIAL COST). *The optimal social cost $Cost_{OPT}$ is the sum of all the players' actual individual costs in case of perfect collaboration.*

An important concept of game-theoretic analysis is the notion of the *Nash equilibrium*. In classic game theory—where there are no Byzantine nodes—this equilibrium describes a situation where no selfish node has an incentive to unilaterally change its strategy. In the following, we extend the definition of a Nash equilibrium to incorporate Byzantine nodes. The *Byzantine Nash equilibrium* (BNE) describes a configuration where no selfish player can reduce his *perceived* cost by changing his strategy, given that the strategies of all other players are fixed.[2]

DEFINITION 3.7 (BYZANTINE NASH EQUILIBRIUM (BNE)). *Let $\overrightarrow{a}[i|x]$ be the strategy vector that is identical to $\overrightarrow{a}$ except for the i-th component a_i which is replaced by x. In a Byzantine Nash equilibrium, no selfish player $i \in \mathcal{S}$ has an incentive to change his strategy if the strategies of all other (selfish and Byzantine) players are fixed, i.e.,*

$$\forall i \in \mathcal{S} : \widehat{cost}_i(\overrightarrow{a}) \leq \widehat{cost}_i(\overrightarrow{a}[i|a'_i]),$$

for every possible strategy a'_i.

[2]Notice that we do not define the Byzantine Nash equilibrium with *actual* individual costs, because they are not known to the players.

While the Byzantine Nash equilibrium must be defined by the *perceived* individual costs, the resulting social cost is determined by the *actual* costs. After all, it is the actual individual costs that players will eventually have to pay. In the following, we will refer to the social cost of the *worst Byzantine Nash Equilibrium* of a problem instance I as $Cost_{BNE}(I, b)$.

It is well-known that selfish and Byzantine players often interact in a manner that yields suboptimal solutions. The degree of degradation resulting from selfish and Byzantine players compared to the social optimum is captured by the *Price of Byzantine Anarchy*.

DEFINITION 3.8 (PRICE OF BYZANTINE ANARCHY). *The Price of Byzantine Anarchy captures how much worse a Byzantine Nash equilibrium can be compared to a collaborative optimal solution. More formally, in a scenario with b Byzantine players, the Price of Byzantine Anarchy $PoB(b)$ is the ratio between the worst-case social cost of a Byzantine Nash equilibrium divided by the minimal social cost, i.e., for all problem instances I,*

$$PoB(b) = \max_I \frac{Cost_{BNE}(I, b)}{Cost_{OPT}(I)}.$$

Note that in the absence of Byzantine players—i.e., if the system consists of selfish players only—, the Price of Byzantine Anarchy is equivalent to the well-known Price of Anarchy (PoA) studied in classic game theory. Specifically, it holds that $PoA = PoB(0)$.

With these definitions, we are ready to define the *Price of Malice* which describes the degree of sub-optimality resulting from malicious Byzantine players in an otherwise selfish system. A high Price of Malice indicates that an economic system is particularly vulnerable to malicious or terrorist attacks. On the other hand, if the Price of Malice is low, the system consisting of selfish players is stable enough to tolerate malicious participants. Clearly, the degree of degradation may depend on the number of Byzantine players in the game. Hence, the Price of Malice is a function of b.

DEFINITION 3.9 (PRICE OF MALICE). *The Price of Malice captures the ratio between the worst Byzantine Nash Equilibrium with b malicious players and the Price of Anarchy in a purely selfish system. Formally,*

$$PoM(b) = \frac{PoB(b)}{PoB(0)}.$$

As we will discuss in detail in Section 4.4, we may also speak of the inverse of the Price of Malice as the game's *Fear Factor* $\Psi(b)$. That is, a game's Fear Factor is given by $\Psi(b) := 1/PoM(b)$.

4. ANALYSIS

In order to derive results for the Price of Malice in various models, we have to establish structural properties of Nash equilibria and the social optimum in the virus inoculation game. We begin with a simple characterization of Nash equilibria if there are no Byzantine nodes. The proof of the following lemma is derived from the analogous lemma in [2].

LEMMA 4.1. *In a pure Nash equilibrium $\overrightarrow{a}$, it holds that (a) every component in the attack graph $G_{\overrightarrow{a}}$ has a size of at most n/L, and (b) inserting any secure node into $G_{\overrightarrow{a}}$ yields a component size of at least n/L.*

PROOF. In the absence of Byzantine players, perceived individual costs equal actual individual costs. (a) Consider a node $u \in S$ in a component of size $t > n/L$. The expected cost is $cost_u(\overrightarrow{a}) = t/n \cdot L > 1$, and hence u could reduce its cost by installing the virus software. This yields a contradiction to the Nash

equilibrium assumption. (b) Consider a secure node u which would be in a component of size $t < n/L$ if it changed to an unvaccinated state. The expected infection cost is smaller than the inoculation cost, i.e., $cost_u(\overrightarrow{a}) = t/n \cdot L < 1$, which also contradicts the equilibrium assumption. □

Lemma 4.1 implies that if $L \geq n$, all nodes will inoculate in the Nash equilibrium. Therefore, for the rest of this paper, we assume that $L < n$.

4.1 Social Optimum

If the inoculation strategies of the individual nodes are planned by a benevolent centralized coordinator, the welfare of the system is maximized. In the following, we derive an asymptotically tight bound on the cost of this social optimum. Throughout this section, perceived costs equal actual costs because when studying the social optimum, we do not consider Byzantine players, i.e., $b = 0$ and therefore $s = n$.

THEOREM 4.2. *The optimal social cost if all players in S act altruistically is $Cost_{OPT} \in \Theta(s^{2/3}L^{1/3})$. More specifically,*

$$\frac{1}{3}\sqrt{\pi} \cdot s^{2/3}L^{1/3} \;\leq\; Cost_{OPT} \;\leq\; 4s^{2/3}L^{1/3}.$$

PROOF. We prove the upper and lower bound in turn.

Lower Bound: If all nodes collaborate to achieve the optimal solution, it holds that $l_i = k_i$ and hence, the social cost is given by

$$Cost \;=\; |I_{\overrightarrow{a}}| + \frac{L}{n}\sum_{i=1}^{l} k_i^2,$$

where $|I_{\overrightarrow{a}}|$ is the number of inoculated nodes, and the k_i's are the sizes of the components in the attack graph. This sum is minimized when all k_i are of equal size, say size K. While each secure node has a cost of 1, every other node has an expected cost of $L \cdot K/n$. Hence, setting $\gamma := |I_{\overrightarrow{a}}|$ and because $s = n$, the optimal social cost can be bounded as

$$Cost_{OPT} \;\geq\; \gamma + (s-\gamma)\left(\frac{LK}{s}\right). \tag{2}$$

A relationship between γ and K follows from a simple geometric argument: If a component in the attack graph is of size K, the number of inoculated nodes at the component's border must be at least $2\pi\sqrt{\frac{K}{\pi}} = 2\sqrt{\pi K}$ (circumference of a disk with volume K). As the total number of such components is at least $\frac{s-\gamma}{K}$ and each inoculated node can be on the border of at most two components, γ can be expressed as

$$\gamma \;\geq\; \frac{s-\gamma}{K} \cdot 2\sqrt{\pi K} \cdot \frac{1}{2} \;=\; (s-\gamma)\sqrt{\frac{\pi}{K}}.$$

By solving this inequality for γ, it follows that $\gamma \;\geq\; s \cdot \frac{\sqrt{\pi/K}}{1+\sqrt{\pi/K}}$. On the other hand, it can be observed that in the optimal solution, for $s > L$, no node is inoculated if all its four neighbors are inoculated. From this, it can be derived that in an optimal solution, $\gamma \leq \frac{s}{2}$. Plugging these two bounds into Inequality (2), the optimal social cost is at least

$$Cost_{OPT} \;\geq\; s \cdot \frac{\sqrt{\pi/K}}{1+\sqrt{\pi/K}} + \frac{LK}{2}.$$

The first term of the above expression is monotonously decreasing in K in the range $0, \ldots, s$, whereas the second one is monotonously increasing. Therefore, taking the minimum of the two terms for a specific K yields a lower bound on $Cost_{OPT}$. When setting

$$K := \frac{2}{3}\sqrt{\pi} \cdot \left(\frac{s}{L}\right)^{2/3},$$

the second term yields $\frac{1}{3}\sqrt{\pi} \cdot s^{2/3}L^{1/3}$. The first term evaluates to $\frac{\sqrt{3/2}\cdot\sqrt[4]{\pi}}{1+\sqrt{3/2}\cdot\sqrt[4]{\pi}} \cdot s^{2/3}L^{1/3} > \frac{1}{3}\sqrt{\pi} \cdot s^{2/3}L^{1/3}$. Consequently, we obtain the following lower bound on the cost of the social optimum:

$$Cost_{OPT} \;\geq\; \frac{1}{3}\sqrt{\pi} \cdot s^{2/3}L^{1/3} \;\in\; \Omega(s^{2/3}L^{1/3}).$$

Upper Bound: Having established a lower bound on the optimal social cost, we now explicitly construct a solution that is asymptotically optimal and proves the tightness of the above lower bound. Given an arbitrary grid $G[r, c]$, we inoculate the nodes as follows. Let $K := (s/L)^{2/3}$. We secure all nodes in the columns $G[\cdot, i\sqrt{K}]$ for $i \in \{1, ..., \lfloor c/(\sqrt{K}+1)\rfloor\}$ and rows $G[i\sqrt{K}, \cdot]$ for $i \in \{1, ..., \lfloor r/(\sqrt{K}+1)\rfloor\}$. Consequently, all attack components are of size at most $\sqrt{K} \times \sqrt{K} = K$ as illustrated in Figure 1 (left). Hence, the total infection cost is at most $L \cdot (s - |I_{\overrightarrow{a}}|)\frac{K}{s} < LK = s^{2/3}L^{1/3}$.

It remains to bound the inoculation cost. In an ideal setting where the components perfectly fit into $G[r, c]$ without leftovers, it holds that for each component of size K in the attack graph, there are exactly $2\sqrt{K}+1$ inoculated nodes. Let X denote the number of components. It holds that $X \cdot (K + 2\sqrt{K} + 1) = s$ and therefore, when plugging in the definition of K, $X = s/[(\frac{s}{L})^{2/3} + 2(\frac{s}{L})^{1/3} + 1]$. The number of inoculated nodes γ is at most

$$\begin{aligned}
\gamma \;&\leq\; X \cdot (2\sqrt{K}+1) \leq \frac{s(2\sqrt{K}+1)}{\left(\frac{s}{L}\right)^{2/3} + 2\left(\frac{s}{L}\right)^{1/3} + 1} \\
&<\; s^{1/3}L^{2/3} \cdot \left(2\left(\frac{s}{L}\right)^{1/3} + 1\right) = 2s^{2/3}L^{1/3} + s^{1/3}L^{2/3} \\
&\leq\; 3s^{2/3}L^{1/3}.
\end{aligned}$$

Combining the infection and inoculation costs, we can bound the optimal social cost by

$$Cost_{OPT} < s^{2/3}L^{1/3} + 3s^{2/3}L^{1/3} = 4s^{2/3}L^{1/3}.$$

□

4.2 Price of Anarchy

The Price of Anarchy compares the social cost of the worst Nash equilibrium (without Byzantine nodes) to the minimal social cost. In the upcoming section, we will first compute $Cost_{NE}$, which is the maximal cost of any Nash equilibrium. Together with the bound for the social optimum in Section 4.1, the Price of Anarchy will follow.

LEMMA 4.3. *The social cost of the worst Nash equilibrium is $Cost_{NE} = \Theta(s)$.*

PROOF. First, we show that $Cost_{NE} = \Omega(s)$. Consider a grid $G[s/L, L]$ consisting of an even number of L rows of size s/L. Assume that columns $G[\cdot, 2i]$ for $i \in \{0, 1, ..., L/2-1\}$ consist of insecure nodes only, while all nodes in the remaining rows are secure. Since all attack components have size s/L, according to Lemma 4.1, this situation constitutes a Nash equilibrium. Observe that every second row is inoculated, engendering an inoculation cost of $s/2$. Moreover, with probability $1/2$, the virus starts at an insecure node, yielding infection cost $s/L \cdot L$. The social cost is therefore $Cost_{NE} = s/2 + 1/2 \cdot s/L \cdot L = s$.

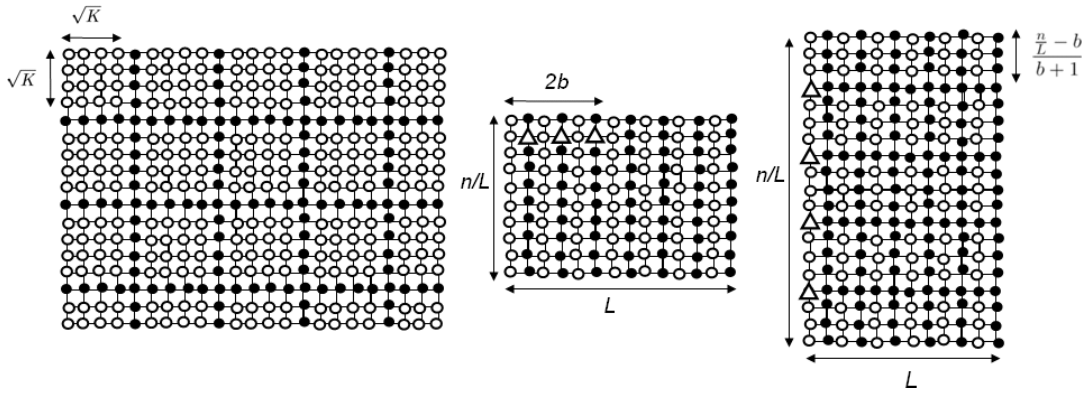

Figure 1: *Left:* Upper bound for social optimum. White nodes are insecure, black nodes are secure. *Middle:* Byzantine Nash equilibrium for $G[n/L, L]$ for the oblivious model. Insecure Byzantine nodes are denoted by white triangles. They are located in a way that may yield an attack component of size $(b+1)n/L+b$. *Right:* Example with large social cost for the non-oblivious, risk-averse model.

It remains to show that $O(s)$ is an upper bound for any Nash equilibrium. Since at most each of the $s = n$ nodes can be inoculated, the inoculation cost cannot exceed s. By Lemma 4.1, we also know that the infected component's size is at most s/L, entailing a total infection cost of at most s as well. Hence, $Cost_{NE} \leq 2s$, and the claim holds. □

By Theorem 4.2 and Lemma 4.3, we get the following result.

THEOREM 4.4. *For the Price of Anarchy (PoA), it holds that*

$$\frac{1}{4} \cdot \left(\frac{s}{L}\right)^{1/3} \leq PoA \leq \frac{6}{\sqrt{\pi}} \cdot \left(\frac{s}{L}\right)^{1/3}$$

PROOF. As for the upper bound, it holds that

$$PoA = \frac{Cost_{NE}}{Cost_{OPT}} \leq \frac{2s}{\frac{1}{3}\sqrt{\pi} \cdot s^{2/3} L^{1/3}} \leq \frac{6s^{1/3}}{\sqrt{\pi} \cdot L^{1/3}}$$

and as for the lower bound, we have $PoA \geq \frac{s}{4 \cdot s^{2/3} L^{1/3}}$. □

4.3 Oblivious Model

We begin our study of the Price of Malice with the oblivious model in which players are clueless about the existence of Byzantine players in the system (cf Section 3). As a consequence, it follows that—since nodes underestimate the attack components' sizes—the nodes' perceived individual costs are smaller than the actual individual costs. It turns out that in the presence of Byzantine nodes, the social costs increase in the number of Byzantine nodes.

LEMMA 4.5. *In the oblivious model, the social cost is at least $Cost_{BNE} \in \Omega(s + \frac{nb^2}{L})$ for $b < \frac{L}{2} - 1$, and $Cost_{BNE} \in \Omega(sL)$ otherwise.*

PROOF. Consider again a grid $G[n/L, L]$ with n/L rows and L columns, where every second column consists of secure nodes only. For simplicity, let L be even. Suppose that in the first b secure columns there is one Byzantine node each, see Figure 1 (middle). In case $b \geq \frac{L}{2} - 1$, every secure column that separates two insecure columns contains one Byzantine node. The remaining Byzantine nodes can be placed at arbitrary places in the secure columns. Because selfish nodes are not aware of the existence of Byzantine nodes in the network, the perceived cost is $\widehat{cost}_i = 1$ for inoculated nodes, and $\widehat{cost}_i = \frac{n/L}{n} \cdot L = 1$ for the other selfish nodes. Hence, the situation constitutes a *Byzantine Nash equilibrium*.

For computing the social costs of this Byzantine Nash equilibrium, we distinguish two cases, depending on whether the number of Byzantine nodes is smaller than $\frac{L}{2} - 1$ or not. For the first case, assume that $b \geq \frac{L}{2} - 1$. Because there is at least one Byzantine node in every secure column that separates two insecure columns has least one Byzantine node, all selfish and Byzantine players form one large attack component. Consequently, each insecure selfish node $i \in S$ is infected with probability 1 and therefore $Cost_{BNE} \geq s \cdot L$.

For the second case, assume that $b < \frac{L}{2} - 1$. Each of the first secure columns contains exactly one Byzantine node. Since L is even, there are $s/2 - b$ secure nodes, and hence the inoculation cost is $s/2 - b$. With probability $((b+1)n/L + b)/n$, the infection starts at an insecure or a Byzantine node of an attack component of size $(b+1) \cdot n/L$, yielding a cost of $(b+1) \cdot n/L \cdot L = n(b+1)$. Moreover, with probability $(s/2 - (b+1)n/L)/n$, an insecure column of size n/L is hit. Thus, for $b < \frac{L}{2} - 1$, we get the following lower bound on the social cost:

$$\begin{aligned} Cost_{BNE} &= \left(\frac{s}{2} - b\right) + \frac{\frac{(b+1)n}{L} + b}{n} \cdot n(b+1) + \\ &\quad + \frac{\frac{s}{2} - (b+1)\frac{n}{L}}{n} \cdot \frac{n}{L} \cdot L \\ &= s + \frac{nb^2}{L} + \frac{nb}{L} + b^2 \in \Omega\left(s + \frac{nb^2}{L}\right). \end{aligned}$$

□

LEMMA 4.6. *In the oblivious model, the social cost is at most $Cost_{BNE} \in O\left(\min\{sL, s + \frac{b^2 n}{L}\}\right)$.*

PROOF. Since at most every selfish node can be inoculated, it is clear that $Cost_{inoc} = O(s)$. It remains to study the infection cost. The infection cost of a node in some component i is L times the probability of this component being hit by the virus, i.e., $L \cdot k_i/n$. Hence, the total infection cost is given by

$$Cost_{inf} = \sum_i l_i \cdot \frac{k_i}{n} \cdot L = \frac{L}{n} \sum_i l_i \cdot k_i,$$

where k_i is the size of the attack components (including Byzantine nodes), and l_i is the number of selfish nodes in this component. In order to upper bound $Cost_{inf}$, let S_{Byz} denote the set of components in the attack graph which contain at least one Byzantine

node, and let $S_{\overline{Byz}}$ be the remaining components. We can rewrite the equation above as

$$Cost_{inf} \;=\; \frac{L}{n} \cdot \left[\sum_{i \in S_{Byz}} l_i \cdot k_i + \sum_{i \in S_{\overline{Byz}}} l_i \cdot k_i \right],$$

that is, we consider the infection cost of components with at least one Byzantine node separately from the remaining "Byzantine-free" components. In the following, let

$$Cost_{inf}^{Byz} := \frac{L}{n} \sum_{i \in S_{Byz}} l_i k_i \qquad Cost_{inf}^{\overline{Byz}} := \frac{L}{n} \sum_{i \in S_{\overline{Byz}}} l_i k_i.$$

We have to prove that neither $Cost_{inf}^{Byz}$ nor $Cost_{inf}^{\overline{Byz}}$ exceeds $O(s + \frac{b^2 n}{L})$.

As we have shown in the proof of Lemma 4.3 in Section 4.2, the total infection cost of a network consisting only of selfish nodes cannot exceed s. Because in our case nodes are oblivious about the existence of Byzantine nodes, attack components without Byzantine nodes behave like in an entirely selfish environment. Therefore, $Cost_{inf}^{\overline{Byz}} \in O(s)$.

It remains to compute the infection cost of those attack components which include at least one Byzantine node. Let b_i be the number of Byzantine nodes in the i-th component in S_{Byz}, and note that $\sum_i b_i = b$. By Lemma 4.1, we know that in the absence of Byzantine nodes, the size of an attack component is at most $k_i \leq n/L$. Therefore, one Byzantine node can increase a component by at most n/L nodes plus itself. From this it follows that the size of an attack component i is bounded by

$$k_i \;\leq\; (b_i + 1) \cdot \frac{n}{L} + b_i, \quad \text{and} \quad l_i \;\leq\; (b_i + 1) \cdot \frac{n}{L}.$$

Using this relationship between b_i and the size of the attack component, we can bound $Cost_{inf}^{Byz}$ as

$$\begin{aligned}
Cost_{inf}^{Byz} &= \frac{L}{n} \sum_{i \in S_{Byz}} l_i \cdot k_i \\
&\leq \frac{L}{n} \sum_{i \in S_{Byz}} \left[(b_i + 1) \cdot \frac{n}{L} \cdot \left((b_i + 1) \cdot \frac{n}{L} + b_i \right) \right] \\
&= \sum_{i \in S_{Byz}} \left[(b_i + 1)^2 \frac{n}{L} + b_i (b_i + 1) \right] \\
&< \sum_{i \in S_{Byz}} \left[(b_i + 1)^2 \left(\frac{n}{L} + 1 \right) \right] \\
&= \left(\frac{n}{L} + 1 \right) \cdot \sum_{i \in S_{Byz}} (b_i + 1)^2.
\end{aligned}$$

Given the constraint that $b_i \geq 1$ for every b_i, and because $\sum_i b_i = b$, the above convex function assumes its maximum for a single positive $b_i = b$. Consequently,

$$\begin{aligned}
Cost_{inf}^{Byz} &\leq \left(\frac{n}{L} + 1 \right) \cdot \sum_{i \in S_{Byz}} (b_i + 1)^2 \\
&\leq \left(\frac{n}{L} + 1 \right) \cdot (b + 1)^2 \;\in\; O\left(\frac{b^2 n}{L} \right).
\end{aligned}$$

On the other hand, it clearly holds that at most every selfish node can be infected and hence, $Cost_{inf}^{\overline{Byz}} + Cost_{inf}^{Byz} \leq sL$. The proof is concluded by adding the upper bounds for $Cost_{inoc}$, $Cost_{inf}^{\overline{Byz}}$, and $Cost_{inf}^{Byz}$. □

Combining Lemmas 4.5 and 4.6 leads to the following theorem that captures the social cost in the virus inoculation game in the presence of b Byzantine players among selfish, oblivious nodes.

THEOREM 4.7. *The social cost in a Byzantine Nash equilibrium with b Byzantine nodes in the oblivious model is $Cost_{BNE} \in \Theta(s + \frac{b^2 n}{L})$, for $b < \frac{L}{2} - 1$, and $Cost_{BNE} \in \Theta(sL)$, otherwise.*

PROOF. In both cases, the lower bound follows from Lemma 4.5. As for the upper bound, note that for $b < \frac{L}{2} - 1$ and due to $L \leq n = s + b$, it holds that $b < \frac{s+b}{2}$ and therefore, $b < s$. Then, the term $s + \frac{b^2 n}{L}$ asymptotically cannot exceed the term sL and therefore, the claim follows. As for the second case, note that for $b \geq \frac{L}{2} - 1$, the term sL is asymptotically smaller or equal to $s + \frac{b^2 n}{L}$. □

Finally, we can derive tight bounds on the The Price of Byzantine Anarchy and the Price of Malice by bringing together the results of Theorems 4.2, 4.4, and 4.7.

THEOREM 4.8. *In the virus inoculation game with b Byzantine nodes among selfish, oblivious nodes, the Price of Byzantine Anarchy and the Price of Malice are*

$$\begin{aligned}
PoB(b) &\in \Theta\left(\left(\frac{s}{L} \right)^{1/3} \left(1 + \frac{b^2}{L} + \frac{b^3}{sL} \right) \right) \quad \textit{and} \\
PoM(b) &\in \Theta\left(1 + \frac{b^2}{L} + \frac{b^3}{sL} \right)
\end{aligned}$$

for $b < \frac{L}{2} - 1$. Otherwise, it holds that

$$PoB(b) \;\in\; \Theta\left(s^{1/3} L^{2/3} \right) \quad \textit{and} \quad PoM(b) \;\in\; \Theta(L).$$

PROOF. Consider the case $b < \frac{L}{2} - 1$. For the Price of Byzantine Anarchy, we have $PoB(b) = \frac{Cost_{BNE}}{Cost_{OPT}} = \frac{\Theta(s + \frac{b^2(b+s)}{L})}{\Theta(s^{2/3} L^{1/3})} \in \Theta\left(\left(\frac{s}{L} \right)^{1/3} \cdot \left(1 + \frac{b^2}{L} + \frac{b^3}{sL} \right) \right)$. From this, the Price of Malice is computed as follows $PoM(b) = \frac{PoB(b)}{PoA} \in \Theta\left(1 + \frac{b^2}{L} + \frac{b^3}{sL} \right)$. The case $b \geq \frac{L}{2} - 1$ follows along the same lines by plugging in the corresponding expressions of Theorem 4.7. □

Our results on the Price of Malice in the oblivious case support the intuition that in the absence of knowledge about the existence of Byzantine players, the quality of the global solution (i.e., the resulting social cost) deteriorates as the number of malicious players increases. In the next section, we will show that the situation may change as soon as selfish players are *aware* of the existence of Byzantine players.

4.4 Non-oblivious Model

Having studied the oblivious model, we now turn our attention to the non-oblivious case in which selfish players are aware of the existence of Byzantine players. If selfish nodes knew about the exact locations of Byzantine nodes, they would be able to compute their optimal choice exactly. If selfish nodes only know the *number of Byzantine nodes* in the system, however, the optimal strategy of a player becomes more complex, and the impact on the social cost more interesting. Specifically, it turns out that in this non-oblivious case, the "Fear Factor" may actually encourage players to act less selfishly and cooperate. Put differently, there may be settings in which the existence of Byzantine players helps to improve the global social cost, rendering the Price of Malice less than 1.

Observe that in the non-oblivious case, every selfish node inoculates if $b \geq \frac{n}{L}$, implying a social cost of s. If $b < \frac{n}{L}$, the resulting social costs are bounded by the following lemma.

LEMMA 4.9. *For $b < \frac{n}{2L}$, the social cost in a Byzantine Nash equilibrium in case of non-oblivious, risk-averse players with b Byzantine nodes is at least*

$$Cost_{BNE} \;\geq\; \frac{s}{2} + \frac{bL}{4}.$$

For all values of b, it holds that $Cost_{BNE} \geq \frac{s}{2}$.

PROOF. We start with the more interesting case $b < \frac{n}{2L}$. Consider a grid with L columns each containing n/L nodes. All nodes in columns $2i+1$ for $i = 0, 1, \ldots, \frac{L}{2} - 1$ and all nodes in rows $j \cdot \frac{n/L-b}{b+1}$ for $j = 1, 2, \ldots, b$ are inoculated. That is, as illustrated in Figure 1 (right), each component of insecure selfish nodes is of size $\frac{n/L-b}{b+1}$.

First, we show that this configuration constitutes a Byzantine Nash equilibrium in the risk-averse, non-oblivious case with b Byzantine nodes. Consider an *insecure node* in some column i. If all b secure nodes in this column are Byzantine, the size of the resulting attack component is $(n/L-b)/(b+1)\cdot(b+1)+b = n/L$. Hence, i's perceived infection cost is

$$\widehat{cost}_i \;=\; L \cdot \frac{(n/L-b)/(b+1)\cdot(b+1)+b}{n} \;=\; 1,$$

which equals the cost of inoculation. Next, consider an *inoculated selfish node* i and distinguish two cases. In the first case, i separates two components consisting of insecure selfish players and a change of i's strategy would merge two components of size $(n/L-b)/(b+1)$ into a single connected component of insecure selfish nodes. Every Byzantine node can connect another component of size $(n/L-b)/(b+1)$ (and itself) to the component containing i. Therefore, the size of the resulting attack component can be as large as $\left(2 \cdot \frac{\frac{n}{L}-b}{b+1} + 1\right) + \left(b \cdot \frac{\frac{n}{L}-b}{b+1} + b\right) \;=\; \frac{b+2}{b+1}\left(\frac{n}{L} - b\right) + b + 1 \;>\; \frac{n}{L} + \frac{1}{b+1}$. The perceived cost of i without inoculation is therefore

$$\widehat{cost}_i \;>\; L \cdot \frac{\frac{n}{L} + \frac{1}{b+1}}{n} \;=\; 1 + \frac{L}{n(b+1)} \;>\; 1.$$

In the second case, we consider a "crossing" node i that is located in the crossing of a secure row and column. Consider the column to the right (or to the left) of i. If all inoculated nodes in this column are Byzantine, the entire column plus node i becomes one large attack component. Hence, the perceived cost of i is

$$\widehat{cost}_i \;>\; L \cdot \frac{\frac{n}{L} + 1}{n} \;>\; 1.$$

In other words, no selfish node has an incentive to change its strategy and the situation in Figure 1 (right) constitutes a Byzantine Nash equilibrium. In the sequel, we lower bound the social cost of this equilibrium under the assumption that all b Byzantine nodes are in column 1. Note that our construction guarantees that this is always possible if $b < \frac{n}{2L}$.

We start with the sum of the infection costs $Cost^0_{inf}$ of insecure nodes in column 0. The number of insecure, selfish nodes in this component is $\frac{n}{L} - b$. Hence, the expected sum of infection costs is

$$Cost^0_{inf} = \left(\frac{n}{L} - b\right) \cdot \frac{\frac{n}{L} - b + b}{n} \cdot L \;=\; \frac{n}{L} - b.$$

Let μ be the number of insecure nodes in columns $3, 5$, etc. The sum of the infection costs $Cost^r_{inf}$ of the remaining attack components (each being of size $\frac{n/L-b}{b+1}$) is

$$Cost^r_{inf} \;=\; \mu \cdot \frac{\frac{n}{L} - b}{n(b+1)} \cdot L \;>\; \mu \cdot \left(\frac{1}{b+1} - \frac{L}{n}\right).$$

Because the number of insecure nodes in these small attack components is $\mu = \frac{L-1}{2} \cdot \left(\frac{n}{L} - b\right)$, it follows that

$$\begin{aligned} Cost^r_{inf} \;&>\; \frac{L-1}{2} \cdot \left(\frac{n}{L} - b\right) \cdot \left(\frac{1}{b+1} - \frac{L}{n}\right) \\ &>\; \frac{1}{2(b+1)}\left(n - \frac{n}{L} - bL + b\right) - \frac{L}{2}. \end{aligned}$$

Finally, we also need to calculate the total inoculation cost of this topology. Clearly, all $s/2$ nodes in even columns are secure. (Recall that column and row indices start with 0.) Furthermore, b nodes in each odd column (except for the first column) are also inoculated. Hence, the total inoculation $Cost_{inoc}$ cost becomes

$$Cost_{inoc} \;=\; \frac{s}{2} + \frac{bL}{2} - b \;=\; \frac{s}{2} + b\left(\frac{L}{2} - 1\right).$$

Combining the various costs, the social cost of the Byzantine Nash equilibrium is

$$\begin{aligned} Cost_{BNE}(b) \;&\geq\; \frac{s}{2} + b\left(\frac{L}{2} - 1\right) + \frac{n}{L} - b \\ &\quad + \frac{1}{2(b+1)}\left(n - \frac{n}{L} - bL + b\right) - \frac{L}{2} \\ &\geq\; \frac{s}{2} + \frac{bL}{4} \end{aligned}$$

for $b \leq \frac{n}{2L}$ and $b \geq 3$.

Finally, note that if $b \geq \frac{n}{2L}$, at least half of the selfish nodes inoculate and hence, $Cost_{BNE}(b) \geq s/2$. □

With this lower bound on the social cost of a Byzantine Nash equilibrium, we can now derive the Price of Byzantine Anarchy as well as the Price of Malice for the non-oblivious, risk-averse model.

THEOREM 4.10. *In the non-oblivious, risk-averse model with b Byzantine nodes, the Price of Byzantine Anarchy is at least*

$$PoB(b) \;\geq\; \frac{1}{8}\left(\left(\frac{s}{L}\right)^{1/3} + \frac{b}{2}\left(\frac{L}{s}\right)^{2/3}\right)$$

for $b < \frac{n}{2L}$. For all b, it holds that $PoB(b) \geq \frac{1}{8}(\frac{s}{L})^{1/3}$.

PROOF. Lemma 4.9 gives us a lower bound on the social cost of a Byzantine Nash equilibrium in the non-oblivious, risk-averse model with b malicious nodes. On the other hand, we have seen in Lemma 4.2, that the optimal social cost is at most $4s^{2/3}L^{1/3}$. Hence,

$$PoB(b) \;\geq\; \frac{\frac{s}{2} + \frac{bL}{4}}{4s^{2/3}L^{1/3}} \;=\; \frac{1}{8}\left(\frac{s^{1/3}}{L^{1/3}} + \frac{bL^{2/3}}{2s^{2/3}}\right).$$

The second lower bound follows analogously. □

THEOREM 4.11. *In the non-oblivious, risk-averse model with b Byzantine nodes, the Price of Malice is*

$$PoM(b) \;\geq\; \frac{\sqrt{\pi}}{48}\left(1 + \frac{bL}{2s}\right)$$

for $b < \frac{n}{2L}$. For all b, it holds that $PoM(b) \geq \frac{\sqrt{\pi}}{48}$.

PROOF. In order to derive the Price of Malice, we can apply our bound from Theorem 4.10 and the upper bound on the Price of Anarchy established in Theorem 4.4. Specifically,

$$PoM(b) \;=\; \frac{PoB(b)}{PoA} \;\geq\; \frac{\frac{1}{8}\left(\left(\frac{s}{L}\right)^{1/3} + \frac{b}{2}\left(\frac{L}{s}\right)^{2/3}\right)}{\frac{6s^{1/3}}{\sqrt{\pi}\cdot L^{1/3}}}.$$

The theorem then follows from arithmetic simplifications. Again, the second lower bound follows in an analogous way. □

Discussion: From a technical point of view, this result shows that the Price of Malice can potentially be less than 1 in the non-oblivious model of the virus inoculation game. Intuitively, it is clear that in the presence of Byzantine players, nodes may be more willing to pay for inoculation. However, we find it interesting that the selfish players' awareness of the existence of malicious Byzantine players may lead to an *improvement* of the overall system behavior, i.e., the *social welfare*. Specifically, the existence (or even the threat!) of malicious Byzantine players can render it worthwhile for nodes to cooperate better.

This highlights the possible existence of a *Fear Factor*, which describes the gain of the overall social efficiency in a selfish system if selfish players are afraid of malicious, Byzantine individuals among them. This Fear Factor is determined by the ratio between the social cost of the worst Byzantine Nash equilibrium with b malicious players and the worst Nash equilibrium in a purely selfish system. Technically, we can define the Fear Factor Ψ as the inverse of the Price of Malice, i.e.,

$$\Psi(b) \ := \ \frac{1}{PoM(b)}.$$

In other words, the Fear Factor Ψ quantifies how much the threat of a common enemy can unite selfish individuals, and to what degree the global social performance is improved.

In the virus inoculation game, the Fear Factor may be both negative and positive. What is interesting to note, however, is that this Fear Factor Ψ cannot be arbitrarily large, regardless of the number of Byzantine players b in the system. Instead, the Price of Malice can never drop below the constant $\frac{\sqrt{\pi}}{48}$ and hence, the Fear Factor is upper-bounded by $\Psi \leq \frac{48}{\sqrt{\pi}}$. That is, the social welfare or efficiency gained due to the Fear Factor cannot exceed a factor of $\Psi \leq \frac{48}{\sqrt{\pi}}$.

The existence of a Fear Factor has been documented in various economic and social models. By combining a game theoretic framework with the classic notion of Byzantine players from distributed computing and cryptography, our model in this paper allows for an *analytical quantification of a system's Fear Factor* Ψ from a computational point of view.

5. STABILITY CONSIDERATIONS

In the previous section, we have studied the degradation of the social welfare in a selfish system caused by Byzantine players. However, besides trying to reduce the optimality of certain outcomes of games, Byzantine players might also attack the *stability* of a system. In this section, we therefore continue our studies by capturing the amount of instability that can be caused by Byzantine players in an otherwise selfish system. Particularly, we are interested in the question, how many Byzantine players suffice in order to keep the system from stabilizing.

In the following, we generalize the model of Section 4 to *arbitrary* network graphs. We assume that the Byzantine players aim at destabilizing the system by repeatedly announcing to have changed from insecure to secure state and back in a worst-case fashion. Thereby, we consider an oblivious model where selfish nodes are not aware of the stability attack. We use the following definitions.

DEFINITION 5.1 (b-STABLE / b-INSTABLE). *We call a game b-stable if b Byzantine players cannot prevent the system from reaching a Nash equilibrium. Similarly, a game is called b-instable if b Byzantine players are sufficient such that no Nash equilibrium will ever be reached in case of oblivious selfish players.*

For the virus inoculation game, the following stability properties can be shown.

THEOREM 5.1. *(i) Generally, the virus inoculation game is not 1-stable. (ii) For certain restricted classes of network graphs, the virus inoculation game is 1-stable. (iii) The virus inoculation game is always 2-instable.*

PROOF. *Claim (i)*: This claim already holds in simple graphs. Assume that n/L is an integer and that $L > 1$, and consider a one-dimensional chain of nodes $\{0, 1, ..., n-1\}$. Let the nodes $i \cdot n/L$ be secure, for $i \in \{0, 1, ..., L-1\}$. By Lemma 4.1, this situation constitutes a Nash equilibrium. Now assume that node n/L is Byzantine, and that it changes to the insecure state. Then, all other nodes $j \in \{1, 2, 3, ..., n/L-1, n/L+1, ..., 2n/L-1\}$ have an incentive to inoculate. However, once such a node j has become secure, node n/L can return to the secure state, yielding components of size smaller than n/L. Consequently, j is bound to become insecure again. These changes can be repeated forever.

Claim (ii): Interestingly, there are robust graphs where no single node can destabilize the system. To see this, consider a complete graph where each node is connected to all other nodes. From Lemma 4.1, it follows that in this network, all Nash equilibria have just one single attack component. Let $\mathcal{C}$ denote the set of nodes of this component, and let $\overline{\mathcal{C}} := V \setminus \mathcal{C}$ be the set of the remaining (secure) nodes. Also by Lemma 4.1, it holds that in any Nash equilibrium, the size of $\mathcal{C}$ is either n/L or $n/L-1$. Moreover, observe that independently of which node is Byzantine and of how the Byzantine node acts, a situation will eventually be reached with the two components as described above. However, the system having converged to such a state, there exist only four possibilities: either the Byzantine node belongs to the node set $\mathcal{C}$ or to the node set $\overline{\mathcal{C}}$, and either $|\mathcal{C}| = n/L$ or $|\mathcal{C}| = n/L-1$. It is run of the mill to verify that in all cases, a Byzantine node can enforce at most one additional change.

Claim (iii): We use the fact that in the virus inoculation game, a pure Nash equilibrium always exists, and that in the absence of Byzantine nodes, selfish nodes stabilize quickly [2]. Assume that the Byzantine nodes first act like selfish nodes until such a classic Nash equilibrium is reached. Now consider an arbitrary secure node $u_1 \in V$, and assume it is Byzantine. If u_1 becomes insecure, according to Lemma 4.1, an attack component $\mathcal{C}$ emerges which consists of n/L or more nodes. If $|\mathcal{C}| > n/L$, at least one node v in $\mathcal{C}$ has an incentive to change to a secure state. Let $\mathcal{C}'$ be the component of v when u_1 is secure, but not v. Assume that after v has changed, u_1 becomes secure again. There are two possibilities. If $|\mathcal{C}'| < n/L$, v will return to insecure state, and the changes can be repeated forever with only one Byzantine node. If $|\mathcal{C}'| = n/L$, a second Byzantine (previously insecure) node u_2 in $\mathcal{C}'$ can force v to become insecure again.

Finally, if $|\mathcal{C}| = n/L$, nodes are indifferent between becoming secure or not. Of course, however, another Byzantine node on the edge of $\mathcal{C}$ can cause endless changes also in this case. □

6. CONCLUSION

What happens when terrorists meet egoists? In this paper, we advocate the study of distributed, potentially economic or social systems consisting of interacting players who can be selfish or Byzantine. Using these models, we have derived bounds on the *Price of Malice* in oblivious and non-oblivious systems. Moreover, we have quantified and upper bounded the *Fear Factor*, which is the *gain* in system efficiency arising from the increased willingness of selfish individuals to cooperate caused by malicious players.

We believe that our paper opens several directions for future research. For example: What is the Price of Malice in a virus inoculation game on other graphs, e.g., on a *small-world graph*? What is the Price of Malice of other games? It seems that while in certain selfish routing games where a single node can attract a lot of traffic by announcing short distances to all other nodes which results in a large Price of Malice, in congestions games the impact of Byzantine players may be much smaller. Another direction for future work is to study the *impact of knowledge* on the resulting Fear Factor in non-oblivious models. Specifically, one could assume that players are not only aware of the existence of Byzantine players, but also of their approximate whereabouts or their statistical distribution. Intuitively, such additional knowledge should decrease the selfish players' incentive for collaboration and thus lower the Fear Factor.

In a larger context, we believe that modeling and studying the notions of *Price of Malice* and *Fear Factor* may lead to new insights in areas beyond those typically found in computer science and networking. Specifically, our framework may be a tool for analytically capturing socio-economic artefacts arising in entirely different fields, including for example political economics or sociology.

7. REFERENCES

[1] A. Aiyer, L. Alvisi, A. Clement, M. Dahlin, J.-P. Martin, and C. Porth. BAR Fault Tolerance for Cooperative Services. In *Proc.* 20^{th} *ACM Symposium on Operating Systems Principles (SOSP)*, pages 45–58, 2005.

[2] J. Aspnes, K. Chang, and A. Yampolskiy. Inoculation Strategies for Victims of Viruses and the Sum-of-Squares Partition Problem. In *Proc.* 16^{th} *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 43–52, 2005.

[3] N. Bailey. The Mathematical Theory of Infectious Diseases and Its Applications. *Hafner Press*, 1975.

[4] D. Dolev. The Byzantine Generals Stike Again. *Journal of Algorithms*, 3(1):14–30, 1982.

[5] S. Eidenbenz, V. Kumar, and S. Zust. Equilibria in Topology Control Games for Ad Hoc Networks. In *Proc. Discrete Algorithms and Methods for Mobile Computing (DIALM-POMC)*, 2003.

[6] K. Eliaz. Fault Tolerant Implementation. *Review of Economic Studies*, 69:589–610, 2002.

[7] A. Fabrikant, A. Luthra, E. Maneva, C. H. Papadimitriou, and S. Shenker. On a Network Creation Game. In *Proc.* 22^{nd} *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 347–351, New York, USA, 2003.

[8] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker. A BGP-based Mechanism for Lowest-Cost Routing. In *Proc.* 21^{st} *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 173–182, 2002.

[9] M. Fischer, N. Lynch, and M. S. Paterson. Impossibility of Distributed Consensus with one Faulty Processor. *Journal of the ACM*, 32(2):374–382, 1985.

[10] C.-Y. Koo. Broadcast in Radio Networks Tolerating Byzantine Adversial Behavior. In *Proc.* 23^{rd} *ACM Symposium on the Principles of Distributed Computing (PODC)*, pages 275–282, 2004.

[11] E. Koutsoupias and C. Papadimitriou. Worst-Case Equilibria. *Lecture Notes in Computer Science*, 1563:404–413, 1999.

[12] L. Lamport, R. Shostak, and M. Pease. The Byzantine Generals Problem. *ACM Trans. Program. Lang. Syst.*, 4(3):382–401, 1982.

[13] D. Malkhi and M. Reiter. Byzantine Quorum Systems. *Journal of Distributed Computing*, 11(4):203–213, 1998.

[14] T. Moscibroda, S. Schmid, and R. Wattenhofer. On the Topologies Created by Selfish Peers. In *Proc.* 25^{th} *ACM Symposium on Principles of Distributed Computing (PODC)*, 2006.

[15] N. Nisan and A. Ronen. Algorithmic Mechanism Design. In *Proc.* 31^{rd} *ACM Symposium on Theory of Computing (STOC)*, pages 129–140, 1999.

[16] M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 2000.

[17] C. Papadimitriou. Algorithms, Games, and the Internet. In *Proc.* 33^{rd} *ACM Symposium on Theory of Computing (STOC)*, pages 749–753, 2001.

[18] R. Pastor-Satorras and A. Vespiagnani. Immunization of Complex Networks. *Physical Review Letter*, 65, 2002.

[19] T. Roughgarden. *Stackelberg Scheduling Strategies*. In *Proc.* 33^{rd} *ACM Symposium on Theory of Computing (STOC)*, pages 104–113, 2001.

[20] T. Roughgarden. *Selfish Routing and the Price of Anarchy*. MIT Press, 2005.

[21] R. Shostak, M. Pease, and L. Lamport. Reaching Agreement in the Presence of Faults. *Journal of the ACM*, 27(2):228–234, 1980.

[22] T. K. Srikant and S. Toueg. Simulating Authenticated Broadcasts to Derive Simple Fault-Tolerant Algorithms. *Journal of Distributed Computing*, 2(2):80–94, 1987.

[23] J. L. Welch and N. Lynch. A New Fault-Tolerant for Clock-Synchronization. *Information and Communication*, 77:1–36, 1988.

[24] A. C. Yao. Protocols for Secure Computations. In *Proc.* 23^{rd} *Symposium on Foundations of Computer Science (FOCS)*, pages 160–164, 1982.