An Investigation of the Vulnerabilities of Scale Invariant Dynamics in Large Teams

Robin Glinton, Paul Scerri, Katia Sycara
Robotics Institute, Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA
rglinton, pscerri, katia@cs.cmu.edu

ABSTRACT

Large heterogeneous teams in a variety of applications must make joint decisions using large volumes of noisy and uncertain data. Often not all team members have access to a sensor, relying instead on information shared by peers to make decisions. These sensors can become permanently corrupted through hardware failure or as a result of the actions of a malicious adversary. Previous work showed that when the trust between agents was tuned to a specific value the resulting dynamics of the system had a property called scale invariance which led to agents reaching highly accurate conclusion with little communication. In this paper we show that these dynamics also leave the system vulnerable to most agents coming to incorrect conclusions as a result of small amounts of anomalous information maliciously injected in the system. We conduct an analysis that shows that the efficiency of scale invariant dynamics is due to the fact that large number of agents can come to correct conclusions when the difference between the percentage of agents holding conflicting opinions is relatively small. Although this allows the system to come to correct conclusions quickly, it also means that it would be easy for an attacker with specific knowledge to tip the balance. We explore different methods for selecting which agents are Byzantine and when attacks are launched informed by the analysis. Our study reveals global system properties that can be used to predict when and where in the network the system is most vulnerable to attack. We use the results of this study to design an algorithm used by agents to effectively attack the network, informed by local estimates of the global properties revealed by our investigation.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

General Terms

Algorithms, Theory, Experimentation

Keywords

Emergent behavior, Self-organisation, Distributed problem solving

1. INTRODUCTION

Cite as: An Investigation of the Vulnerabilities of Scale Invariant Dynamics in Large Teams, R. Glinton, P. Scerri and K. Sycara, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 677-684.

[1] In the near future, large heterogeneous teams of robots, agents, and people will be utilized to solve problems in a variety of applications including search and rescue and the military. The sheer size of such teams will mean that the amount of data collected by the team will be overwhelming for its constituents. For this reason, team members will need to share concise information abstractions to maintain shared situational awareness.

The physics of communication, along with environmental constraints, will require team members to communicate via a point to point associates network. This will in turn lead to complex information dynamics and emergent phenomena, which in turn leads to unpredictability.

This paper shows that small amounts of anomolous information introduced to such a belief sharing system can cause errors on a system-wide scale due to the intrinsic dynamics of the system. This could potentially be exploited by a malicious agent attempting to disrupt such a system. Both analytical and empirical evidence is provided to support this assertion.

Previous attempts to describe the vulnerabilities of complex networked system primarily focus on finding vulnerabilities in the network topology without consideration of the dynamics of the process taking place on the network [1]. In this work, the dynamics on the network have a dramatic impact on the vulnerability of the system. Studies which have considered how to influence network dynamics of a complex system include [2]. These all focus on a single type of information spread whereas here we can have conflicting data that fundamentally changes the dynamics and introduces new vulnerabilities due to the way information is fused on the network.

It was recently shown that a team of agents could tune their local trust such that the frequency distribution of cascades of changes in belief followed a power law [3]. When the team was tuned like this, the team's ability to rapidly reach correct conclusions despite noisy data and limited communications was shown to be dramatically higher. However, in this paper we show that when a system is tuned like that, it also becomes extremely vulnerable to malicious attack.

We conduct an analysis to show that for a system exhibiting scale invariant dynamics, a single anomalous sensor reading could result in a number of agents on the order of the size of the system coming to the incorrect conclusion. The analysis compares the rate at which the probability that an agent is on the edge of coming to a correct conclusion, called the percolation probability, increases relative to the same probability for an incorrect conclusion. The anal-

[1] This research has been sponsored in part by AFOSR FA95500810356.

ysis reveals that these two numbers remain close until the agents in the system converge. Although this difference is biased towards correct conclusions, the analysis shows that this difference is small enough for a few anomalous sensor readings to push large numbers of agents towards incorrect conclusions.

To confirm the predictions of the analysis we empirically explore the effect of injecting a single incorrect sensor reading into the system on the correctness of conclusions reached by agents in the system. We show empirically by exhaustively searching trajectories of system execution that there is always a point in that trajectory where injecting a single sensor reading can lead to system wide incorrect conclusions. We further show that an adversary could mount an effective attack on the system if the adversary had global knowledge of the distance of the system from the percolation threshold for the incorrect conclusion.

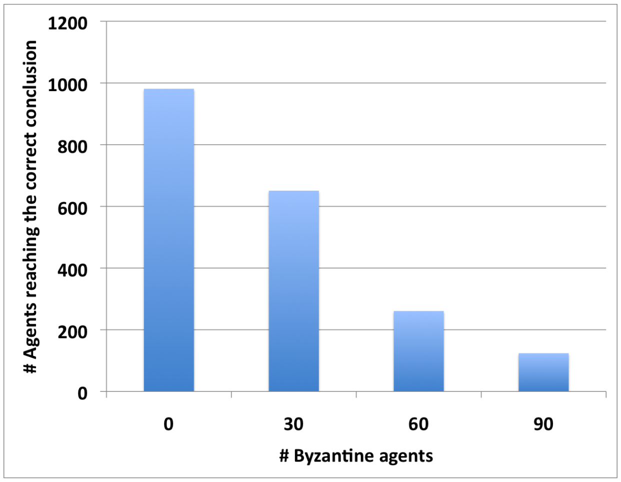

Figure 1: Belief sharing system exhibiting scale invariant dynamics is vulnerable to a small percentage of Byzantine agents.

Just as complex systems can be attacked from external sources, it is also possible for attacks to originate from within. Thus it is necessary to understand the potential vulnerabilities of such a system to threats from within. To this end we study the vulnerability of the agents within the system to reaching incorrect conclusions as a result of the action of Byzantine agents within the system. Specifically, we study mechanisms for picking the most vulnerable points in the network for attack by Byzantine agents. We explore Several different mechanisms for selecting which nodes are Byzantine, using methods typically employed in the study of the vulnerabilities in network topologies to network disintegration. The study reveals that the most effective method is that which selects the nodes with the maximum number of neighbors. Finally, our study shows that as the number of Byzantine agents in the network increases, the trust range between agents that results in a scale invariant distribution of cascades is no longer optimal. As the number of byzantine agents increases the optimal value of trust is lowered slightly with the agents becoming slightly more conservative to account for the misinformation circulating in the system.

In a large distributed system it is unlikely that an adversary would have access to the global network state or topology, thus it is desirable to study whether an effective attack on the system could be launched using only local knowledge of the network state and topology. To investigate the feasibility of a practical attack we developed a local algorithm, inspired by [4], where Byzantine agents use knowledge of the local connectivity and a local estimate of the percolation threshold to decide when and where to focus an attack. We found that such an attack is as effective, in reducing the number of agents that come to a correct conclusion, as an attack mounted with full knowledge of the system state and network topology.

The remainder of this paper is organized as follows: Section 2 gives an overview of the model of a belief sharing system used to study emergent vulnerabilities. Section 3 presents an analysis that reveals a vulnerability of such a system to small amounts of anomalous information introduced by an adversary. Section 4 empirically explores the vulnerability of the system to spoofed sensor readings introduced by an adversary with global system knowledge. Section 5 empirically explores the vulnerability of the system to Byzantine agents with detailed knowledge of the network topology and state. Section 6 explores the feasibility of effective attacks based on partial knowledge of the system. Section 7 presents the related work and Section 8 presents conclusions and future work.

2. MODEL

In this section, we formally describe the underlying model used in the remainder of the paper. A cooperative team of agents, $A = \{a_1, \ldots, a_{|A|}\}$ are connected by a network, $G = (A, E)$ where E is the set of links in G which connect the agents in A. An agent a_i may only communicate directly with another agent $a_j \in N_{a_i}$ if $\exists e_{i,j} \in E$ where we refer to the set N_{a_i} as its *neighbors*. The average number of neighbors that the agents in G have is defined as $< d >$ where $< d >= \frac{\sum_i |N_{a_i}|}{|A|}$.

Sensors, $S = \{s_1, \ldots, s_{|S|}\}$ provide noisy observations to the team. Only one agent can directly see the output of each sensor. The sensors return binary observations about some fact b from the set $\{$*true*,*false*$\}$. We refer to the probability that a sensor s will return a correct observation as its reliability r_s. The reliability of a sensor is known to the agent that receives observations from it.

In the remainder of this paper, unless otherwise specified, $|A| = 1000$, $|S| = |A|/20$ and $r_s = 0.55 \forall s$.

A key assumption of the model is that it is infeasible for agents to communicate actual sensor observations to one another and that they may only communicate whether they currently believe the fact to be *true*, *false* or if they are undecided, *unknown*.

Each agent a_i uses either an observation received from a sensor or conclusions about b communicated by neighbors to form a belief $P_{a_i}(b \rightarrow true)$ about b. A new observation is incorporated into the current belief to form a new belief $P'_{a_i}(b \rightarrow true)$ using an expression of Bayes' Rule with *cp* as the conditional probability that the neighbors conclusion is correct. In this model *cp* acts as a measure of the trust between agents.

An agent will come to a conclusion about the truth of the fact and communicate this conclusion to neighbors if its belief in that conclusion exceeds a fixed threshold. The details of the belief update calculation and thresholding were taken from [3]. When an agent comes to a conclusion and communicates with neighbors, the neighbors may then come to a conclusion and communicate. This chain of conclusion formation is called a *cascade*. Previous work showed that agent conclusions are most accurate when the probability $P(c)$ that c agents change their belief during a cascade is given by $P(c) \propto c^{-3/2}$. The most important metric used in this paper is T_a, the number of agents in the network coming to the correct conclusion. We define the system under study to be vulnerable if there are small sets of Byzantine agents $\hat{a}$, subset $\hat{a} \subset |A|$ such that T_a and $|\hat{a}|$ is greatly reduced. Another objective is to find times at which the system is most vulnerable to the injection of anomalous sensor readings and agent conclusions.

3. THEORY OF SCALE INVARIANT VULNERABILITY

In this section we conduct an analysis that reveals a vulnerability to attack in systems which exhibit scale invariant dynamics. Such systems have been shown to enable very accurate and efficient belief fusion using very little communication. We would like to leverage this efficiency to design practical information fusion systems. However, first it is necessary to understand potential vulnerabilities of such a system to adversarial action. To this end we analyze the difference in the rate at which agents in the system reach correct conclusion (called the percolation probability for that conclusion) relative to the rate at which they approach incorrect conclusions. Our, analysis reveals that although agents overwhelmingly reach correct conclusions at a higher rate, the difference to the rate at which they near incorrect conclusions is small. This suggest that when the majority of agents are close to making a decision, a single anomalous sensor reading could offset this balance causing a large percentage of agents to reach the incorrect conclusion instead.

Previous work [5] showed that the probability of a large cascade disseminating a conclusion system wide is given by:

$$\sum_{\hat{k_t}} \sum_{\hat{s_t}} \beta(\hat{k_t}, \beta(\hat{g_t})) P(\hat{k_t} \mid \hat{s_t}) P(\hat{g_t} \mid \hat{s_t}) \beta(\hat{s_t}) \qquad (1)$$

This occurs when the percolation probability for that conclusion exceeds a threshold called the percolation threshold.

The vector $\hat{k_t} = [k_0, k_1, \ldots, k_t]$ gives the sequence of the sizes of avalanches that occurred over time. Similarly the vector $\hat{g_t}$ gives the sequence of false avalanches that occurred. (Note for the model presented in this paper, only a single cascade per time step is possible). Finally the vector $\hat{s_t} = [s_0, s_1, \ldots, s_t]$ gives the sequence of sensor readings input to the system up until time t. The terms in Equation 1 are as follows: The term $\beta(\hat{s_t})$ gives the probability of a specific sequence of sensor readings input to the system, $P(\hat{k_t} \mid \hat{s_t})$ and $P(\hat{g_t} \mid \hat{s_t})$ give the probability of a resulting sequence of cascade sizes of correct and incorrect conclusions respectively. Finally $\beta(\hat{g_t}, \hat{k_t})$ gives the probability that a random agent in the network will be touched by a net number of correct cascades such that it is one correct communication from a neighbor away from reaching the correct conclusion.

Starting with Equation 1 we show that the difference between the probability of a large cascade of correct conclusions and a large cascade of incorrect conclusions is small just before a large cascade of correct conclusions occurs, revealing a vulnerability in the system. To facilitate ease of computation we simplify Equation 1 by observing that the scale invariant distribution is heavy tailed, meaning that the probability of a cascade of size 1 is close to 1. It is then reasonable to assume that before a large cascade occurs, all cascades are of size 1. Under this assumption, given a specific sequence of sensor readings $\hat{s_t}$, all of the probability mass of the cascade sequence distribution $P(\hat{k_t} \mid \hat{s_t})$ collapses to a single possible sequence of cascades. With this simplification Equation 1 is reduced to Equation 2:

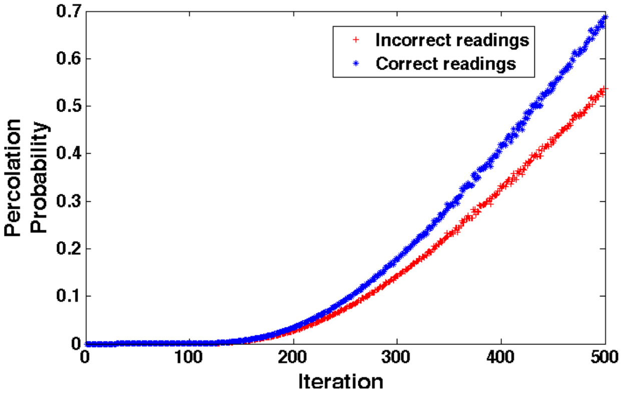

Figure 2: The percolation probability for incorrect information stays near that for correct information until just before the threshold is exceeded.

$$\sum_{\hat{k_t}} \sum_{\hat{s_t}} \beta(\hat{k_t}, \hat{g_t}) \beta(\hat{s_t}) \qquad (2)$$

All three terms of this equation are binomially distributed. We can further simplify computation using this expression by recognizing that a binomial distribution can be approximated by a normal distribution. The first term in the equation which gives the probability of the difference between the number of competing cascades that reached the agent is then normally distributed with mean $\mu = \frac{n_T - n_F}{|A|}$ and standard deviation $\sigma = \frac{n_T}{|A|}\frac{1}{1-|A|} + \frac{n_T}{|A|}\frac{1}{1-|A|}$. Where n_T and n_F give the number of correct sensor readings and incorrect sensor readings in the the sequence $\hat{s_t}$. The probability of a net number of false cascades touching the agent is obtained by simply switching the n_T and the n_F in the normal distribution.

With this substitution it is easy to numerically integrate Equation 2, to give the percolation probability for correct and incorrect cascades. The result of this computation is shown in Figure 2. The x-axis of this figure gives the timestep and the y-axis gives the percolation probability. For the random network the calculation was conducted for, the percolation threshold that would result in a large cascade is .33. In the figure it is evident that the percolation probability for a correct cascade reaches this threshold first. However, at this point the percolation probability for the large incorrect cascade is 0.25. This difference corresponds to less than 5% of the agents in the system. For the system under study with $|A| = 1000$, this is less than 50 agents. This estimate is a maximum because the analysis was predicated on only avalanches of size 1 occuring and the assumption of a loop free network. In practice relaxing either of these conditions would reduce the number of agents necessary to upset the balance and cause a large cascade of incorrect information.

The conclusion is that relatively few sensor readings or a small number of Byzantine agents could potentially cause a system on the verge of large numbers of agents reaching the correct conclusion to have the exact opposite occur. Furthermore, this result suggests that the system is particularly vulnerable near the percolation threshold. Although the curves in Figure 2 are even closer together at lower percolation probabilities, additional Byzantine agents or anomalous sensor readings would be required to drive the system closer to the percolation threshold. For example at iteration 300 an additional 150 agents would need to be influenced to drive the system to the percolation probability for a cascade of incorrect conclusions. Figure 4 illustrates the vulnerability. As the agents in the system near a correct conclusion, there is a smaller group of agents on the edge of an incorrect conclusion. A single anomalous sensor correction can set off a large cascade among such agents leading to large numbers of agents coming to incorrect conclusions.

4. BYZANTINE SENSORS

In this section we investigate the vulnerability of a system exhibiting scale invariant dynamics of belief exchange to small amounts of anomalous sensor readings introduced to sensors by a malicious attacker. The results of this section show that for all of the network topologies with the exception of Small World, there was always a point in time at which injecting a spurious sensor reading would result in large numbers of agents reaching the incorrect conclusions. In addition, experiments reveal that an adversary with knowledge of the number of agents in the system 1 communication from a neighbor away from reaching a correct conclusion could use this information to decide the best times to introduce spurious sensor readings into the network for maximum impact on the conclusions reached by agents with minimal intervention.

We conducted experiments to explore this potential system vulnerability. In the first experiment we test if an adversary with total knowledge of the system, including all of the possible trajectories of the system dynamics could cause the agents in the system to adopt the wrong conclusions. In this experiment we exhaustively search trajectories of the system simulation for points where introducing a single incorrect sensor reading will result in a cascade for which greater than half of the agents in the system incorporate the incorrect sensor reading into their belief. The exact procedure for searching the system trajectories is as follows. First, a *snapshot* of the system is taken where the current belief state of all of the agents is recorded. Next, we exhaustively consider what would happen if incorrect sensor readings were introduced to every permutation of two sensors in the system. For each such permutation, the resulting cascades, if any, are allowed to propagate until the system quiesces. The agents are then restored to their states before the introduction of the incorrect readings before the next permutation is explored. If a large cascade does not result, the agents are returned to their previous belief state and the system is allowed to evolve as if the intervention did not occurr. The entire procedure is then repeated.

In this experiment we recorded the number of large cascades that occurred as a result of malicious intervention during 10 rounds of the above procedure, where each round consists of 100 steps, where each step consists of the permutation search discussed above. The parameter values used during this experiment were $|A| = 1000$, $|S| = 1/20|A|$, $s_r = 0.55$, and $< d >= 4$. The results of the experiment are given by Figure 3. The x-axis gives cp and the y-axis gives the number of rounds out of 10 in which greater than 50% of the agents incorporated the incorrect information artificially introduced to the sensors.

The plot shows that during almost every round, there is a point in the system trajectory where introducing incorrect information at the sensors would have resulted in a large cascade, propagating this incorrect information to more than half the agents in the system. This only occurs for rounds when the value of cp approaches the value which results in scale invariant dynamics. This suggests that an omniscient agent could almost always cause the agents in the system to come to the incorrect conclusion. This of course is not practical and in the next experiment we investigate what information could be used by a malicious actor to mount a practical attack on the system.

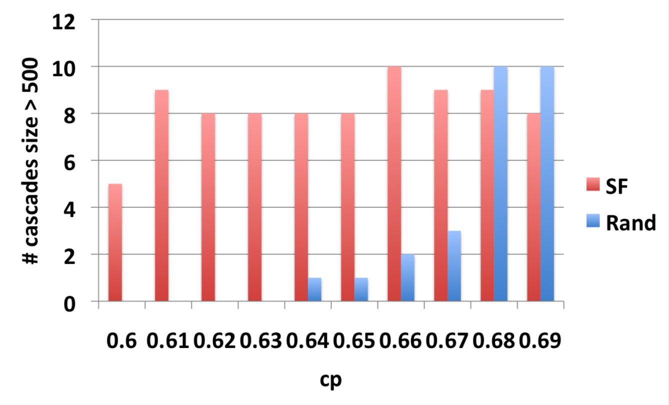

Figure 3: Cascades resulting from malicious intervention at sensors.

The preceding experiment showed us that there is almost always a point in the trajectory of the system where the system is extremely vulnerable to malicious intervention using a small amount of misinformation. However, the experiment did not tell us anything about *when* the system is most vulnerable. Specifically, the experiment did not reveal what properties of the system could be used by a malicious actor to decide when to inject misinformation at the sensors. We hypothesize, due to the results of Section 3 that the system would be most vulnerable to such an attack when the system is on the edge of making a decision. That is when the agents are approaching a percolation threshold for a large correct avalanche.

The percolation threshold in this case is a network specific probability that a randomly selected agent requires a single communication from a neighbor to come to a conclusion. We conducted an experiment to test this hypothesis. In this experiment we simulated 1000 runs of the system and injected a single incorrect reading at a randomly chosen sensor using two methods to decide when to inject the reading. In the first method we simply randomly selected the time-step at which to inject the incorrect sensor reading. In the second method the reading was injected when the percolation probability was at the percolation threshold. We repeated this for each network topology under study. The results are given in Table 5.

Network	Random success rate	Percolation success rate
SF	0%	95%
R	0%	63%
SW	0.03%	83%

Figure 5: The effect of using the percolation probability of the network to decide when to attack the network compared to random attack.

5. BYZANTINE AGENTS

In this section we investigate the vulnerability of a belief sharing system exhibiting scale invariant dynamics to attacks on, or malfunction of the agents that exchange fused information within the system. We experiment with three methods for selecting Byzantine nodes, all based on global knowledge of the network topology.

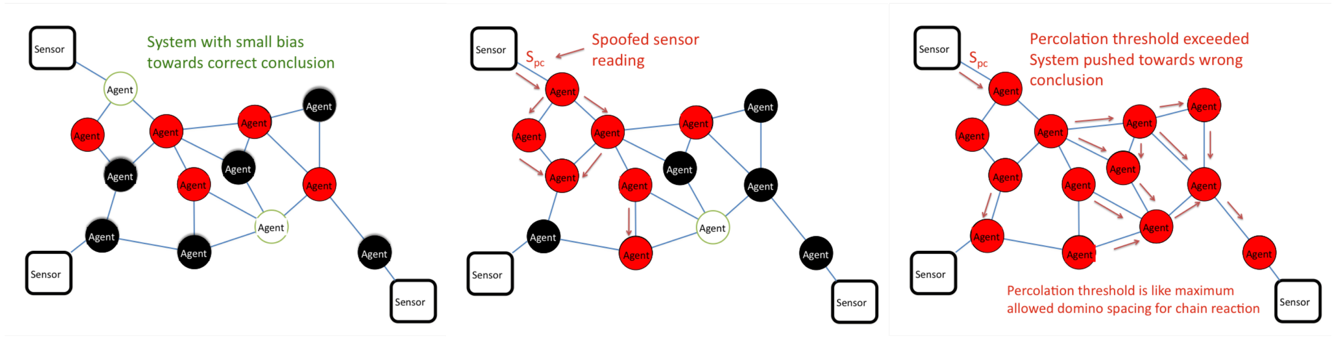
Sensor
System with small bias
towards correct conclusion
Agent
Sensor
Sensor
Sensor
S_{pc}
Spoofed sensor
reading
Sensor
Sensor
Percolation threshold exceeded
System pushed towards wrong
conclusion
Percolation threshold is like maximum
allowed domino spacing for chain reaction

Figure 4: A single incorrect sensor reading introduced by an adversary can percolate through the network causing widespread incorrect conclusions.

The analysis of Section 3 revealed that changing the decisions of a relatively small number of agents in the system could dramatically reduce the number of agents reaching the correct conlusion. This suggests that a small number of Byzantine agents could influence the conclusions of the majority of the agents in the system by sharing incorrect information or noise.

In this section we analyze this assertion by empirically exploring the vulnerability of the system to the action of malicious or malfunctioning agents. Specifically we analyze the effect of malicious agents in the system on the performance of the system as measured by the the number of agents in the system that reach the correct conclusion. We investigated two types of Byzantine agents. The first type of Byzantine agents we investigated pathologically share incorrect information. The second type shares random information. Both types of agent simply ignore any information received from neighbors or sensors.

One of the key results of this section is that a relatively small number of Byzantine agents dramatically reduces the number of agents that reach a correct conclusion over all network types. In addition, we find that the trust value that results in scale invariant dynamics is no longer optimal when a small number of Byzantine agents are present.

5.1 Byzantine agent selection

Three different methods of selecting which nodes in the simulation would be Byzantine were used in experiments. In the first method, nodes are simply drawn at random from a uniform distribution over all of the nodes in the system. The second method which we call the maximum influence method is a modified version of the method due to Kleinberg [6]. Using this method, a node is selected by the number of nodes that would be *infected* by a cascade starting at that node. We call this a nodes influence number. The nodes with the highest influence numbers are selected. To calculate the influence number of node Q each node is initially marked uninfected. Next node Q communicates with its neighbors. When these neighbors receive this communication they draw a real number in the range $[0, 1]$ from a uniform distribution. If this number exceeds a threshold, the node marks itself and communicates with neighbors otherwise it does nothing. When all communication ceases, the influence number of Q is the number of nodes in the network marked infected.

The third method of Byzantine node selection picks the nodes with the largest number of neighbors, this is called the max node method. Figure 6 shows the node that would most likely be selected first in a particular graph structure. The max node method picks the node that simply has the highest fanout while the max influence method is biased towards the node with the most pathways to the other nodes in the network.

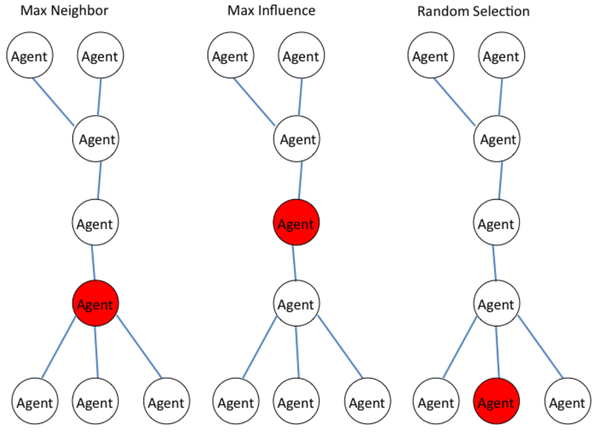

Figure 6: Three methods used for selecting Byzantine nodes.

5.2 Byzantine agent experiments

First we conducted an experiment to investigate the result on system performance of Byzantine nodes which pathologically share incorrect information with neighbors. In the experiment, we investigated how system performance as measured by the number of agents reaching the correct conclusion, changed as the number of Byzantine nodes in the system was varied. Experimental parameters are as follows: The number of Byzantine nodes in the system was varied from 0-10% of $|A|$ in increments of 1%. All remaining graphs in this section were produced using the parameter values $|A| = 1000$, $|S| = 50$, $r_c = 0.2$, and $r_s = 0.55$, and $< d >= 4$. We also varied the structure of the communication network used by the agents. The results are given by Figures 7 and 8. In Figure 7 the x-axis gives the number of Byzantine agents out of 1000. The y-axis gives x the number of agents out of 1000 that come to the correct conclusion. Each curve represents a different network topology including Random, Scale Free, and Small Worlds networks. The leftmost plot shows the results when Byzantine nodes are selected at random, the middle plot shows the results for nodes selected using the maximum influence method, and the leftmost plot shows the results when nodes that have the largest number of neighbors are selected.

The first trend evident across all of the communication networks is that with a relatively small percentage of Byzantine nodes, the number of agents that comes to the correct conclusion drops dramatically. In fact with 10% of the nodes in the system, only the Scale-Free network has greater than half the agents in the network reaching the correct conclusion. The theoretical limit, due to Lamport [7], says that agents in a network can reach a correct consensus with a maximum of 33% of the nodes as Byzantine. The system under study requires less than 10% of the agents to be Byzantine, to prevent a correct consensus in the truth of the fact being monitored. Also all networks are most vulnerable when nodes with the maximum number of neighbors are chosen.

For the ScaleFree network, the trend of the vulnerability of this system with respect to the way nodes are selected for the injection of misinformation, reflects the known results for the vulnerability of the ScaleFree network to the removal of nodes. The ScaleFree network proves most robust of all of the networks when Byzantine nodes are selected at random, with 60% of the agents reaching the correct conclusion with 10% of the nodes Byzantine. Conversely, the ScaleFree network is most vulnerable when the nodes with the largest number of neighbors are selected. In this case, with only 1% of the nodes Byzantine, less than 10% of the agents in the network come to the correct conclusion. This can be explained by the extremely skewed distribution of the number of neighbors that each nodes has in a scale free network. A Scale Free network has a long tailed distribution, with a few nodes, called hubs, having a large number of neighbors and the remainder of the nodes having relatively few neighbors. When nodes are selected at random, their is a low likelihood that the hubs will be selected and the fused information originating at the hubs overwhelms that spread by the Byzantine nodes. Conversely, the hubs have a disproportionately large effect on the network spreading misinformation widely when they are Byzantine.

The second trend evident across all of the networks is that for the Random network topology, there is a distinct threshold in the number of Byzantine nodes beyond which the number of agents that reach the correct conclusion drops suddenly and dramatically. This threshold is 6% of the agents for both the random agent selection and selection for agents with the maximum number of neighbors. This threshold drops to 4% when the maximum influence method is used for node selection.

All network topologies perform about the same for the maximum influence method of selecting nodes to be Byzantine. This suggests that the specific dynamics of this system have a large effect on the which nodes are vulnerable within the system. Otherwise the generic influence spreading, which is dependent on the static topology of the network itself, would be much more effective at means of picking Byzantine nodes to cause the maximum number of nodes to come to the incorrect conclusion.

The network with the Small Worlds topology shows a linear drop in the number of agents reaching the correct conclusion with increasing numbers of Byzantine agents.

Over all for a system with these dynamics, the Scale Free network topology would be the best choice. It is least vulnerable to all attacks except attacks on the hubs. Since the hubs in the network are relatively few, they would take a relatively small amount of computational resources within a system to monitor for intrusion. Furthermore, an attacker would need a large amount of information about the topology of the network to select nodes for attack effectively. Below its vulnerability threshold, the random network is the least vulnerable, and would be the best choice of network topology in a secure environment where an attacker could only select relatively few nodes to attack.

Figure 8 shows how the value of cp which results in the largest number of agents reaching the correct conclusion, and hence which associated system dynamics as discussed in Section 2 are least vulnerable, as the number of Byzantine nodes in the system changes. The x-axis of the figure gives the number of Byzantine agents out of 1000 in the system. The y-axis gives the center of mass of cp. The center of mass is the mean value of cp over simulation runs weighted by the number of agents that reach the correct conclusion for that value of cp. The center of mass is defined mathematically as $\sum_i \frac{cp_i * nT_i}{nT_i}$, where nT_i is the number of agents that reached the correct conclusion for simulation run i. The most notable trend in Figure 8 is that for the network with the Random topology there is a distinct shift of the trust value cp that gives the best performance, away from the value that results in scale invariant dynamics.

The high level conclusion of this Section is that the *scale invariant* dynamics that were previously showed to lead to high accuracy in the conclusions of agents, leaves the system vulnerable to intervention by a small number of Byzantine nodes. This means that a system utilizing scale invariant dynamics, or that intrinsically had such dynamics would either need to operate in a very secure environment, or explicitly have a mechanism to detect the presence of Byzantine nodes.

6. ATTACKS WITH LIMITED INFORMATION

In previous sections experiments have shown that an adversary could dramatically reduce the accuracy of agent's conclusions using global system knowledge. However, in practice, it is more likely that an adversary would have only partial knowledge of the system. To investigate the vulnerability of the system to attacks based on partial system knowledge, we developed an algorithm used by Byzantine agents to attack the system using only local information about the system. In sections4 and 5 we found that the system was most vulnerable at times when close to the percolation threshold in agent decisions. We also found that most networks exhibiting scale invariant dynamics were most vulnerable at nodes with many neighbors. For this reason, the Byzantine agents executing our attack strategy use local estimates of the percolation threshold in the network to decide when to attack and knowledge of the local network topology to decide where to attack.

The details of the algorithm are as follows. The Byzantine agent draws a random number in the range $[1, |A|]$ from a uniform probability distribution. If this number falls below a threshold, which we call the activity threshold, the agent continues to operate normally, fusing conclusions of neighbors and communicating the resulting conclusion. This threshold is intended to ensure that only a preselected percentage of the Byzantine agents in the system are active at any time. If the random number drawn by the agent exceeds the activity threshold, the agent estimates the distance of the agent and it's neighbors from the percolation threshold. The knowledge of the percolation threshold suggests that the attacker would have knowledge of the high level topology of the network (e.g. Random vs. Scale Free) but not specific details of the connectivity in the network. If this estimate is within a given distance from the

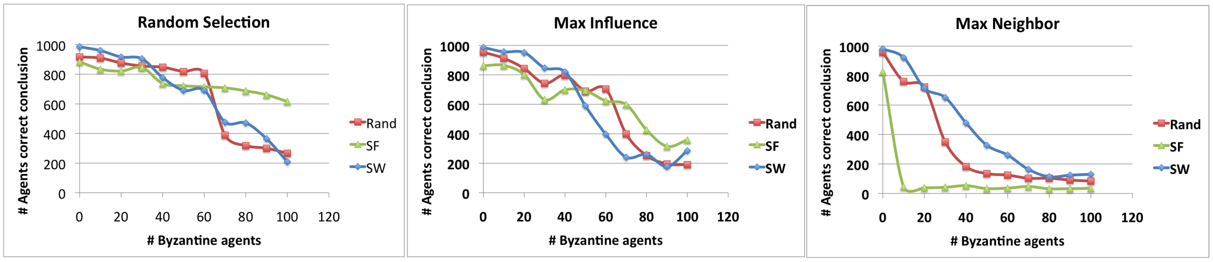

Figure 7: The effect of Byzantine nodes on the correctness of the conclusions of agents across the three methods for selecting Byzantine nodes.

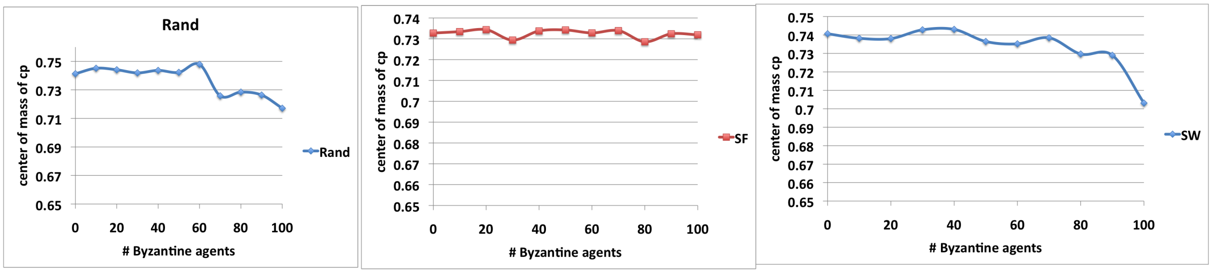

Figure 8: The effect of Byzantine nodes on the cp that gives best performance.

percolation threshold for the network, the agent then sends several incorrect conclusions to its neighbor that has the highest number of network links.

We conducted an experiment to test the efficacy of this algorithm over a range of networks. The parameters used are the same as those for previous sections. The activity threshold is varied between 0.05 and 0.30 (effectively varying the number of active Byzantine agents between 5% and 30%. This is plotted against the average number of agents that reach the correct conclusion. This plot is shown in Figure 9. The plot shows that relatively few agents

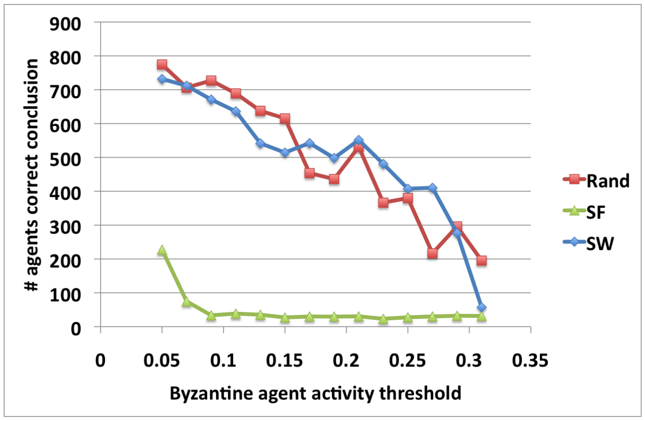

Figure 9: The effect of Byzantine nodes using only local knowledge of the system on the accuracy of the conclusions reached by agents in the network.

using the algorithm, dramatically reduce the number of agents in the system reaching correct conclusions over a range of network topologies.

7. RELATED WORK

There have been several studies conducted to investigate models whose dynamics are governed by cascades on complex networks. These include models of fads[8, 9], rumors [10], gossip[11], forest fires [12], and diseases[13, 14]. Common to all of these models is that the dynamics are governed by the spreading of a single influence. In contrast, our model investigates competing influences which significantly alters the dynamics of a system.

In [15], Parunak presents a model of the collective convergence of agents to a cognitive state. This model is similar to ours in that it does include multiple states that agents can converge to and hence competition between states. Parunak focuses on studying the macroscopic performance of the system. We build upon Parunak's investigation by analyzing the dynamics of the system directly and investigating the relationship between the dynamics and the performance of the system.

A number of studies have investigated the impact of Byzantine nodes on the performance of a distributed system and mechanisms for coping with their presence [16],[17],[18]. We extend these studies by investigating how the efficacy of Byzantine agents are impacted by the dynamics of a system exhibiting scale invariance in belief exchange.

Previous work has extensively explored methods for picking network nodes that are most vulnerable to fracturing the structure of the network [1], [19]. This paper considers the impact of many of the metrics discussed in this body of work on information dynamics on a network by using them for the placement of agents that spread misinformation on a belief sharing network.

Recently there has been significant interest in social networks [20], [21] and the impact of those networks on performance of a group. For example, Xu looked at the impact of networks on routing information to a specific agent [22]. Kleinberg, looked at the impact of the network on the performance of decentralized search algorithms [6], when a single agent has information valuable to the system. We build on both of these contributions by investigating the case when a large percentage of the agents in the team are both sources and sinks for information, which fundamentally changes the dynamics of information exchange in the system.

8. CONCLUSIONS AND FUTURE WORK

When information exchange between agents exhibits scale invariant dynamics, the speed and reliability with which the team can converge to correct conclusions, despite noisy data and highly limited communication is dramatically increased. Before, this property can be leveraged to design efficient information fusion, we need to understand the vulnerability of the system to malicious intervention. In this work we found that scale invariant dynamics make a system susceptible to the presence of Byzantine agents and sensors. We showed analytically that when the agents in the system are near to a correct conclusion, they are simultaneously near to coming to an incorrect conclusion. This leaves the system vulnerable to small amounts of anomalous information and small number of Byzantine agents. We found that Byzantine agents were most effective at reducing the accuracy of the conclusions of other agents when placed at high degree nodes in the network. We further found that attacks were most effective when launched when the network is close to a percolation threshold in the decisions of agents. In future work, we propose to extend the model to capture additional features of information sharing, including beliefs of several variables and a richer communication model, while maintaining the mathematical simplicity that allows the types of detailed analysis above. We also intend to simulate features that are harder to model mathematically, such as the ways mobile sensors might be redeployed based on initial conclusions and how other coordination activities can influence belief convergence. Finally we intend to develop mechanisms for detecting Byzantine or malfunctioning agents and mitigating their impact on system performance informed by the algorithm described in this work.

9. REFERENCES

[1] Y. Chen, G. Paul, R. Cohen, S. Havlin, S. P. Borgatti, F. Liljeros, and H. E. Stanley, "Percolation theory applied to measures of fragmentation in social networks," *Phys. Rev. E*, vol. 75, no. 4, p. 046107, Apr 2007.

[2] M. Lelarge, "Efficient control of epidemics over random networks," in *SIGMETRICS/Performance*, 2009.

[3] R. Glinton, P. Scerri, and K. Sycara, "Exploiting scale invariant dynamics for efficient information propagation in large teams," in *Proc. of AAMAS'10*, 2010.

[4] R. Cohen, S. Havlin, and ben Avraham D., "Efficient immunization strategies for computer networks and populations," *Physical Review Letters*, 2003.

[5] P. Glinton, R.and Scerri and K. Sycara, "An explanation for the efÞciency of scale invariant dynamics of information fusion in large teams," in *Proceedings of Fusion*, 2010.

[6] J. Kleinberg, "Complex networks and decentralized search algorithms," in *Proceedings of the International Congress of Mathematicians (ICM)*, 2006.

[7] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Trans. Program. Lang. Syst.*, vol. 4, no. 3, pp. 382–401, 1982.

[8] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, vol. 100, no. 5, p. 992, 1992. [Online]. Available: http://www.journals.uchicago.edu/doi/abs/10.1086/261849

[9] W. DJ, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Science*, vol. 99, pp. 5766–5771, 2002.

[10] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili, "Theory of rumour spreading in complex social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 374, no. 1, pp. 457–470, 2007.

[11] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Trans. Netw.*, vol. 14, no. SI, pp. 2508–2530, 2006.

[12] S. Clar, B. Drossel, and F. Schwabl, "Scaling laws and simulation results for the self-organized critical forest-fire model," *Phys Rev E*, vol. 50, p. 1009Ð1018, 1994.

[13] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys. Rev. Lett.*, vol. 86, no. 14, pp. 3200–3203, Apr 2001.

[14] V. M. Eguíluz and K. Klemm, "Epidemic threshold in structured scale-free networks," *Phys. Rev. Lett.*, vol. 89, no. 10, p. 108701, Aug 2002.

[15] V. Parunak, "A mathematical analysis of collective cognitive convergence," in *AAMAS '09*, 2009, pp. 473–480.

[16] D. Dolev and H. R. Strong, "Authenticated algorithms for byzantine agreement," *SIAM Journal on Computing*, vol. 12, no. 4, pp. 656–666, 1983.

[17] J.-P. Martin, "Fast byzantine consensus," *IEEE Trans. Dependable Secur. Comput.*, vol. 3, no. 3, pp. 202–215, 2006.

[18] P. Brutch and C. Ko, "Challenges in intrusion detection for wireless ad-hoc networks," *Applications and the Internet Workshops, IEEE/IPSJ International Symposium on*, vol. 0, p. 368, 2003.

[19] H. J. . A.-L. B. RŐka Albert, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, July 2000.

[20] D. Watts and S. Strogatz, "Collective dynamics of small world networks," *Nature*, vol. 393, pp. 440–442, 1998.

[21] A.-L. Barabasi and E. Bonabeau, "Scale free networks," *Scientific American*, pp. 60–69, May 2003.

[22] Y. Xu, P. Scerri, B. Yu, S. Okamoto, M. Lewis, and K. Sycara, "An integrated token-based algorithm for scalable coordination," in *AAMAS'05*, 2005.