Mean Field Games for Continuous Time Density Dependent Markov Chains

by

Dheeraj Narasimha

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2021 by the
Graduate Supervisory Committee:

Lei Ying, Co-Chair
Gautam Dasarathy , Co-Chair
Yongmin Liu
Srinivas Shakkottai

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

The seminal work of Lasry and Lion showed the existence of Nash equilibria in the continuum limit of agents who try to optimize their own utility functions. However, a lot of work in this region is predicated on strong assumptions on the asymptotic independence of the agents and their homogeneity. This work explores the existence of Equilibria under the limit for Markov Decision Processes for density dependent continuous time Markov chains. Under suitable conditions it is possible to show that the empirical measure of the agents converges in finite time to a time invariant distribution which makes the solution of the MDP tractable. This key step allows one to show not only the existence of equilibria for these MDPs without asymptotic independence but also a tractable means to find said equilibria. Finally, this work shows that a fixed point does exist in the infinite state limit. However, to show that such a limit is indeed a Nash equilibrium remains an open problem.

DEDICATION

Dedicated to amma and appa.

ACKNOWLEDGMENTS

Most importantly I would like to thank Prof Ying for his support and guidance throughout this process. I am sure (being the imbecile I am) that this was an arduous process, I will be ever grateful for the guidance, patience and support. I would like to take this chance to thank the J.N. Tata and Dorabji Tata grant for supporting me as I made my journey here. I would like to thank Prof Srinivas, Prof Liu and Prof Dasarathy for their guidance and discussions. I must also take this chance to thank the mentor who started me on this path Prof Rahul Vaze at TIFR (another teacher who bore the brunt of my idiocy) back when I was an infant undergrad who didn't even understand what a random variable was.

Thank you my friends, there are too many to list here but my roommate Parth (who should have jettisoned me out the door a day into meeting me), my lab mates, Xin, Weichang, Hairi, Honghao, and friends outside the lab Suhas and Brandon deserve special mention here. Thank you all for not giving into your impulses and punching me in the head every time you met me.

TABLE OF CONTENTS

APPENDIX Page

Chapter 1

INTRODUCTION

The original setting by Lasry and Lions (2007) for mean field games considered a system with N agents who wished to optimize identical cost functionals over a time horizon in a one-shot game. The devices must choose a control policy that minimizes their cost given their motion depends on this control policy. The caveat in this case is that the cost functional depends on the empirical measure of the agents (which in general is a time varying quantity), as a result the policy itself depends on the empirical measure of the agents. The paper then looked at the problem in the continuum limit of the agents to try and solve the problem of the coupled set of equations (the equation for optimal control and the equation for the evolution of the empirical distribution over time).

It should be noted here that the coupled system requires a two pronged solution. In order to solve the control problem one typically uses the Hamilton Jacobi Bellman equation. The boundary condition for this set of partial differential equations is given in terms of the final time, that is; at the end of the time horizon I have a well defined cost functional. For example if you were trying to decide how a car should move to get to its destination within a certain time horizon (say an hour) you can assign a cost at the end of the hour based on how far the car is from its destination. Therefore, this set of PDEs must be solved backward in time. On the other hand suppose we can describe the motion of the car using the velocity at each point in time. We can now use the Fokker Plank equations to describe the empirical measure as it changes over time. In this case, we know where all the cars are at time 0, so we know our initial empirical measure, thus, we must solve this portion of the problem forward in time.

For this coupled forward-backward problem, the authors wish to know if a Nash equilibrium exists. Lasry and Lions (2007) shows that the answer is in fact to the positive under suitable conditions. We will outline a more thorough (but not even close to complete) description of the more general Mean Field Game model in Appendix C. It should be noted that while the existence of these Nash equilibria was solved in the paper, this does not help us find the NE policy.

A more recent line of work considered repeated games under a variety of different applications, some of these authors include, Iyer *et al.* (2014); Li *et al.* (2016); Xia *et al.* (2018). While the state space in these settings are still coupled and often arbitrary (could be continuous or discrete), the time horizon is often discrete or a jump process. One of the major advantages of such a process is the tractability of the solution since it now becomes possible to circumvent the more complicated Fokker-Plank equations. Examples include games with strategic complementarities and second price auction games. It should be noted here that be it the first setting or the second, the authors assumed asymptotic independence either directly or due to previously known results on the *propagation of chaos*.

As the title suggests, we will look at an N agent setup where each agent's state evolves as a Continuous time density dependent markov process. Each agent wishes to maximize their own utility function which can be described using a Markov decision process, in the simplest case, this process simply averages the utility over time and can be given by a very simple concave functional to be maximized (2 and 4) but in the more general case can be a countably infinite state MDP as in 3. We will lay out some assumptions here that will follow throughout the paper,

Assumption 1:

We assume that the agents are homogeneous. That is, the rates of moving from one state to the next remains constant. Therefore, the only difference between two devices (say i and j) with regards to its state changes will be due to differences in policies ($\alpha^{(i)}$ and $\alpha^{(j)}$).

Our first major step is approximating the empirical distribution. A recent approach based on Stein's method Ying (2016, 2017); Gast (2017); Gast and Van Houdt (2018) can directly establish the convergence of steady-state distributions to the MFE without the interchange of the limits argument and provide the rate of convergence. Using Ying (2016) we show that the empirical measure weakly converges to a time invariant distribution given by the fixed point of an ODE constructed by detailed balance equation.

This means that in the limit as the number of agents grows very large, the fraction of agents in any given state is constant. Note, while this seems intuitive at first glance, this is not straightforward as it first seems. The conventional law of large numbers result does not hold here since the states of the agents are not independent. We call this limit our *mean field limit.* Given the mean field limit each device now has a very simple problem it must solve. In this way the mean field limit greatly simplifies the complexity of the problem.

For each choice of policy chosen by all the agents, we can now find the fixed point of the ODE, let $\{\pi_0, \pi_1, \pi_2, ...\}$ be this fixed point,we now have a map from the policy space to the fraction of monitored regions,

$$T_1 : \alpha \to \{\pi_0, \pi_1, \pi_2, ...\} \tag{1.1}$$

Assumption 2:

Two time scale separation. We assume that there are two time scales of operation.

The map T_1 converges in the *fast time scale*. Once the fraction has reached the steady state, the agents will change their policy to optimize their cost function. This portion occurs on the *slow time scale*. Thus, given the value of $\{\pi_0, \pi_1, \pi_2, ...\}$ each agent will choose a new policy that optimizes their utility. Due to assumption 1, this means that each UAV will choose the same policy, we denote this map from the fraction of monitored regions to the policy space by,

$$T_2 : \{\pi_0, \pi_1, \pi_2, ...\} \to \alpha \tag{1.2}$$

We are now ready to define a mean field equilibrium (MFE). Consider the composition of map T_1 and T_2, given by $T_2 \circ T_1$. This maps from the policy space to the policy space. We call a point α^* a mean field equilibrium if,

$$\alpha^* = T_2 \circ T_1(\alpha^*) = T_2(T_1(\alpha^*)) \tag{1.3}$$

Here for the sake of completeness we define the well known concept of Nash equilibrium. A strategy profile is called a Nash equilibrium $\alpha := \{\alpha^1, \alpha^2, \alpha^3 ... \alpha^N\}$ if for any player i, if everyone else fixes their strategy profile, player i cannot benefit by choosing a different profile, i.e,

$$J(\alpha^i|\alpha^1, \alpha^2, ...\alpha^{i-1}, \alpha^{i+1}...\alpha^N) \geq J(\hat{\alpha}|\alpha^1, \alpha^2, ...\alpha^{i-1}, \alpha^{i+1}...\alpha^N) \tag{1.4}$$

for any valid policy $\hat{\alpha}$.

Remark 1.

- The fixed point is usually proven using Schauder or Brower's fixed point theorem by showing the continuity of the maps from policy space to policy space.

- If we know that the ODE corresponding to map T_1 is globally asymptotically stable and locally exponentially stable then the other conditions to show the validity of map T_1 are simple to verify.

- Given the points above one can actually say that in the finite state space, every MFE is in fact a Nash equilibrium.

- Although this form of generality seems nice, in our experience of proving these results the hardest part of the problem was actually proving the continuity of the maps T_1 and T_2 depending on the case when no closed form expression was available for the maps. And finding and proving that a function is a Lyapunov function to prove the stability results mentioned above.

1.1 Main Results

- **Mean Field Game analysis for distributed MAC** *Age of Information (AoI)* seeks to ensure that the samples received at an aggregation point satisfy constraints on the difference between the current time and generation time of the last received sample, known as "age" Kaul *et al.* (2012); Yates and Kaul (2018); Kam *et al.* (2018); Sun *et al.* (2019). We wish to model this problem of a MAC protocol in terms of a mean field game where the density dependence comes through the fraction of channels that are currently used by the other agents. We show the existence of a mean field limit and regions depending on the parameter where the mean field nash equilibrium exists. Surprisingly there are regions where no MFNE exists at all and the system simply oscillates between points. Finally, as a metric for performance we bound the price of anarchy showing that the difference between a central controller forcing a uniform policy among the agents and our own distributed policy.

- **Age Dependent MAC** This can be viewed as a complete generalization of the previous chapter. Here we use an exponential clock to track the delay over time and use it to come up with a delay sensitive MAC protocol using the mean field game setting. We solve this problem by truncating the system using a space of policies that choose to do nothing beyond some finite state K and show that under these constraints at least one Nash equilibrium exists and perhaps more importantly, computed. In this chapter, we show the existence of fixed points (Mean field equilibria) however, showing that these equilibria are Nash equilibria remains an open problem.

- **Strategic Deconfliction** Lest someone accuse us of using contrived MAC protocols as examples for density - dependent continuous time markov chains for our models we come up with a contrived task allocation problem for UAVs during disaster relief. We are able to show similar results to chapter 2 in this case as well.

Chapter 2

A MEAN FIELD GAME ANALYSIS OF DISTRIBUTED MAC

2.1 System Model and an M-player Game

We consider a multi-channel ultra-dense wireless networks with N channels and $M = mN$ devices as shown in Figure 2.1. At each time instance, one and only one device can transmit over a given channel due to interference. As in many IoT applications, each device wants to continuously communicate their latest status to corresponding receivers, which could be an access point or another IoT device. The messages are called status messages in this report. We note after a new status message is generated, the device does not need to transmit old, unsent status messages currently in the buffer, so the old status messages will be discarded. This communication model is an example where the system wants the most fresh information and wants to minimize the "age of information" Kaul *et al.* (2011).

We assume for each device, status messages are generated according to a Poisson process with rate λ. When the device is probing an idle channel to transmit, it only stores the latest status message. If the device is transmitting a status message when a new status message arrives, the device keeps the newest status message in the buffer and transmits it immediately after finishing sending the one in transmission. A channel being used to transmit a status message is in busy state, otherwise the channel is in idle state. We further assume that the time it takes to transmit a message is exponentially distributed with mean one.

When a device has a status message to transmit, it searches for an idle channel to transmit the message. A device cannot afford to continuously monitor all N frequency

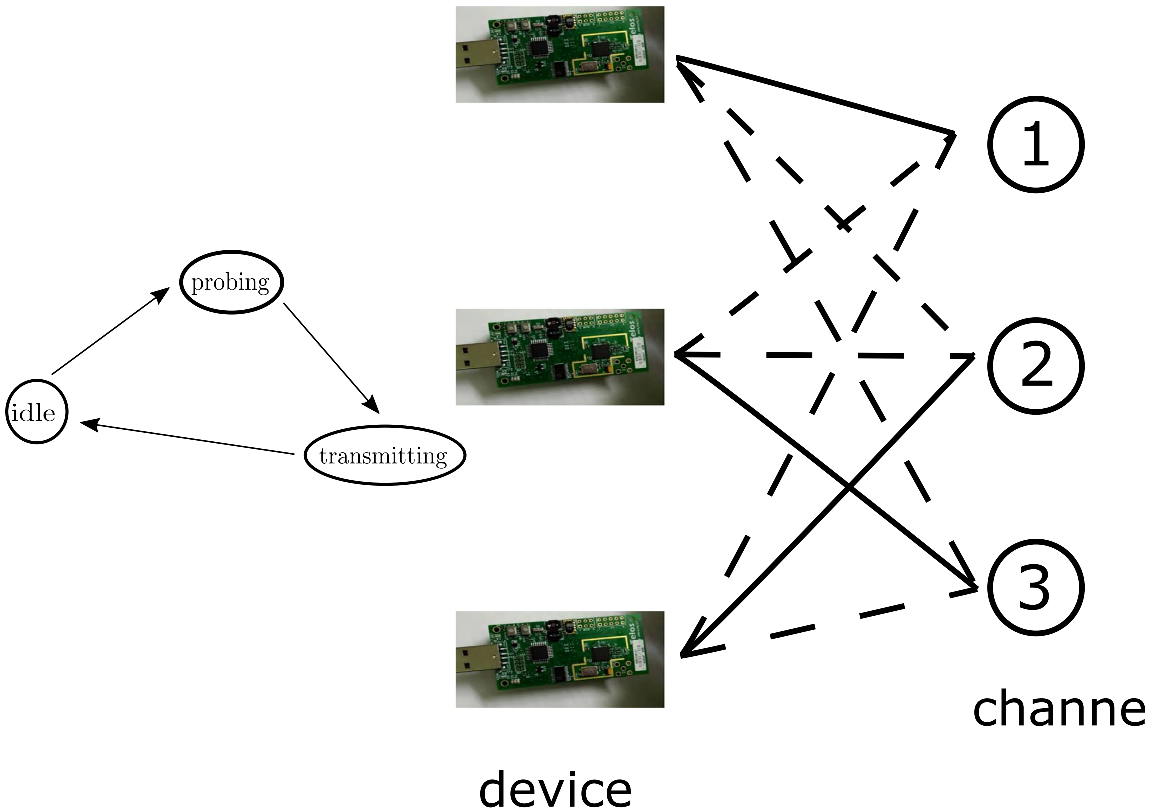


Figure 2.1: A system with $M = 3$ devices and $N = 3$ channels. Each device is a three-state Markov chain.

bands at all times, because channel probing costs energy and battery powered smart wireless devices are energy constrained. We assume each device maintains an internal exponential clock with rate k. When the exponential clock ticks, the device probes $\frac{d}{k}$ channels. If one of the $\frac{d}{k}$ channels is idle, the device occupies the channel and transmits the message in the buffer. A device has three possible states: *idle (0)*, *probing (1)* and *transmitting (2)*. Let $Q_i(t)$ denote the number of devices in state i at time t. Each device is associated with a *continuous-time* Markov chain with three states as shown in Figure 2.2 in principle. The Markov-chain includes three states and the transitions occur as follows:

- The state moves from idle to probing when a message arrives, which occurs with rate λ.

- Let d_l and k_l denote the probing parameters used by device l, and $\mathbf{d}$ and $\mathbf{k}$ denote M-dimensional vectors that represent the probing parameters of all M devices. Given $Q_2(t)$, the number of devices in the transmitting state, by probing $\frac{d_l}{k_l}$ channels, the probability of finding an idle channel is

$$1 - \left(\frac{Q_2(t)}{N}\right)^{\frac{d_l}{k_l}}.$$

  Here we use sampling with replacement to derive the expression, note, as the number of channels go to infinity this is equivalent to sampling with replacement. Therefore, the state of the Markov chain transits from probing to transmitting with rate

$$k_l \left(1 - \left(\frac{Q_2(t)}{N}\right)^{\frac{d_l}{k_l}}\right).$$

- The state transits from transmitting to idle when (1) the status message is transmitted, which occurs with rate one, and (2) no new status message arrives during the transmission, which occurs with probability $\frac{1}{1+\lambda}$. To see this let T denote the transmission time of a message, which is an exponential random variable with mean one. Under the Poisson arrival, the probability of no arrival during a period of duration t is $e^{-\lambda t}$. Therefore, the probability that there is no new message arrival during the transmission is

$$\begin{aligned}
&\Pr(\text{no arrival during transmission})\\
=&E\left[\Pr(\text{no arrival during duration } T|T)\right]\\
=&\int_{t=0}^{\infty} e^{-\lambda t} e^{-t}\, dt\\
=&\int_{t=0}^{\infty} e^{-(\lambda+1)t}\, dt\\
=&\frac{1}{1+\lambda}.
\end{aligned}$$

  Therefore, the transition rate is $\frac{1}{1+\lambda}$.

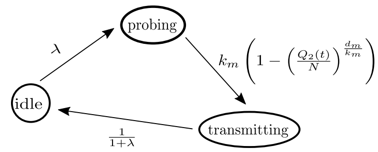


Figure 2.2: The Continuous-Time Markov Chain

Suppose $Q_2(t)$ is a constant, then the stationary distribution of this three-state Markov chain, denoted by π, can be calculated using the global balance equations:

$$\lambda\pi_0 = k_l\left(1-\left(\frac{Q_2}{N}\right)^{\frac{d_l}{k_l}}\right)\pi_1 = \frac{1}{1+\lambda}\pi_2,$$

from which, we have

$$\begin{aligned}\pi_0 &= \frac{1}{\lambda(1+\lambda)}\pi_2\\ \pi_1 &= \frac{1}{(1+\lambda)k_l\left(1-\left(\frac{Q_2}{N}\right)^{\frac{d_l}{k_l}}\right)}\pi_2\\ \pi_2 &= \frac{1}{1+\frac{1}{\lambda(1+\lambda)}+\frac{1}{(1+\lambda)k_l\left(1-\left(\frac{Q_2}{N}\right)^{\frac{d_l}{k_l}}\right)}}.\end{aligned} \tag{2.1}$$

However, $Q_2(t)$ is a random process whose stationary distribution is determined by $\mathbf{d}$ and $\mathbf{k}$ so is difficult to calculate. Now let $\pi^{(l)}(\mathbf{d},\mathbf{k})$ denote the stationary distribution of the Markov chain associated with device l. As mentioned earlier, calculation of $\pi^{(l)}$ is difficult even for fixed $\mathbf{k}$ and $\mathbf{d}$.

Making the problem even more difficult, each device needs to balance the energy consumed for probing and the amount of information transmitted. We consider the following cost function for each device:

$$\hat{J}(d_l,k_l) = -\pi_2^{(l)}(\mathbf{d},\mathbf{k}) + c\left(\pi_1^{(l)}(\mathbf{d},\mathbf{k})d_l\right)^2. \tag{2.2}$$

In the equation above, the first term $\pi_2^{(l)}(\mathbf{d},\mathbf{k})$ is the fraction of time the device is in the transmitting state, so can be viewed as the average throughput. The amount of energy consumed during the transmission of a message is proportional to the size of the message, so the transmission energy is linearly proportional to the throughput and we can view the throughput term also includes the energy consumption for transmissions. In the second term, $\pi_1^{(l)}(\mathbf{d},\mathbf{k})$ is the fraction of time the device is in the probing state and d_l is the number of channels it probes per unit time when it is in the probing state, so $\pi_1^{(l)}(\mathbf{d},\mathbf{k})d_l$ is the average number of channels probed per unit time. c is a constant. The quadratic form is in keeping with the idea that energy usage for a given task is convex for most communication applications. Given other devices' probing parameters $\mathbf{d}_{-l}$ and $\mathbf{k}_{-l}$, device l aims at finding the optimal d_l^* and k_l^* such that

$$\begin{aligned}(d_l^*, k_l^*) &\in \arg\min_{d_l,k_l} \hat{J}(d_l,k_l) \\ &= \arg\min_{d_l,k_l} -\pi_2^{(l)}(\mathbf{d},\mathbf{k}) + c\left(\pi_1^{(l)}(\mathbf{d},\mathbf{k})d_l\right)^2 .\end{aligned} \tag{2.3}$$

We note that this is an M-player game and the difficulty in solving the Nash equilibrium of this M-player game is in calculating $\pi^{(l)}(\mathbf{d},\mathbf{k})$ as discussed earlier.

2.2 Mean-Field Game for Ultra-Dense Wireless Networks

Since solving the M-player game (2.3) is difficult, we use the MFG approach with $N, M \to \infty$. In the next section, we will show that assuming all devices use the same probing policy (d,k), then as $N, M \to \infty$, $Q_i(\infty)/M$ converges weakly to q_i^*, which is the equilibrium point of the following mean-field model:

$$\begin{aligned}\frac{dq_0}{dt} &= -\lambda q_0 + \frac{1}{1+\lambda}q_2 \\ \frac{dq_1}{dt} &= \lambda q_0 - k(1-(mq_2)^{d/k})q_1 \\ \frac{dq_2}{dt} &= k\left(1-(mq_2)^{d/k}\right)q_1 - \frac{1}{1+\lambda}q_2\end{aligned} . \tag{2.4}$$

We defer the derivation of this mean-field model and the proof of convergence to the Appendix A.1. Intuitively, $q_i(t)$ is an approximation of $Q_i(t)/M$ and q_i^* is an approximation of $Q_i(\infty)/M$ at the mean-field limit.

Given q_2^*, the fraction of devices are in transmitting state, the fraction of busy channels is $\gamma^* = mq_2^*$. Now to introduce the MFG, we assume time-scale separation such that devices adapt their probing strategies in a slower time scale than the convergence of the mean-field model. Under this assumption, when it is the time for devices to adapt their probing policies, all devices can measure γ, which can be done accurately under the time-scale separation assumption. Then after measuring the fraction of busy channels is γ, each device can compute the stationary distribution of its three-state Markov chain according to (2.1) by substituting $\gamma = Q_2/N$, and also the corresponding cost $J(d, k)$. Each device optimizes its probing strategy (d^*, k^*) such that

$$(d^*, k^*) \in \arg\min_{d,k} J(d, k), \tag{2.5}$$

where

$$\begin{aligned} J(d,k) = &- \frac{1}{1 + \frac{1}{\lambda(1+\lambda)} + \frac{1}{(1+\lambda)k\left(1-\gamma^{\frac{d}{k}}\right)}} \\ &+ c\left(\frac{d}{(1+\lambda)k\left(1-\gamma^{\frac{d}{k}}\right) + \frac{k\left(1-\gamma^{\frac{d}{k}}\right)}{\lambda} + 1}\right)^2. \end{aligned} \tag{2.6}$$

In other words, choosing a probing strategy to minimize its cost for given γ. Note that the cost function $J(d, k)$ is different from $\hat{J}(d, k)$ defined in (2.2) because γ is a constant in $J(d, k)$ but it is a function of (d, k) in $\hat{J}(d, k)$. We can view $\hat{J}(d, k)$ as the true cost function and $J(d, k)$ is an estimate of the true cost obtained by assuming γ does not change even when the device changes its probing strategy. We use different notations to emphasize the difference.

In summary, given (d, k), the mean-field model (2.4) maps (d, k) to the fraction of busy channels γ. Let T_1 denote this mapping, i.e.

$$T_1 : (d, k) \to \gamma.$$

Given the fraction of busy channels γ, each device minimizes the cost function J in (d, k), which maps γ to policy (d, k). Let T_2 denote this mapping, i.e.

$$T_2 : \gamma \to (d, k).$$

With the notation defined above, we formally define the MFG and Mean Field Nash Equilibrium (MFNE).

MFG for Distributed MAC:

- **Initialization:** All devices are initialized with a common probing policy (d, k).
- **System Adaptation:** The mean-field model (2.4) converges under policy (d, k) and the fraction of busy channels converges to a constant γ.
- **Policy Optimization:** All devices learn γ in the system adaptation step, and optimize their probing strategies by minimizing $J(d, k)$. Go to the system adaptation step. □

A policy (d^*, k^*) is called the MFNE if

$$(d^*, k^*) = T_2(T_1(d^*, k^*)).$$

At the MFNE where all devices use the policy (d^*, k^*), no device has incentive to unilaterally change the strategy in the mean-field limit. We also remark that the assumption that all devices use the same policy (d, k) at the beginning is not critical. Under the assumption all devices have the same cost function, the optimal probing strategy is determined only by γ. Therefore, even devices have different probing strategies at the beginning, after they measure γ in the policy optimization step, they will start to use the same probing policy.

In the next section, we prove the weak convergence of $Q_i(\infty)/M$ to q_i^*, which is the key assumption we have used to derive the MFG.

2.3 Mean-Field Limit with Fixed (d, k)

Assume all devices have the same cost function. Then given the fraction of busy channels γ, the solution of the optimal policy (d^*, k^*) is the same for all devices. Therefore, without loss of generality, we assume all devices use the same policy (d, k) and consider the convergence of the fraction of busy channels to its mean-field limit in this homogeneous case. Before proving this result, we first present the following lemma.

Lemma 1. *The cost function $J(k, d)$ satisfies for any $k < d$,*

$$J(d, d) < J(d, k).$$

Proof. Given γ, k and d, the stationary distribution of the three-state Markov chain is given by (2.1) with $Q_2/N = \gamma$. The cost function $J(k, d)$, therefore, can be written in terms of γ, k, and d as

$$\begin{aligned} J(k, d) = &- \frac{(1+\lambda)k(1-\gamma^{d/k})}{(1+k(1-\gamma^{d/k})(1+\lambda+\frac{1}{\lambda}))} \\ &+ c\left(\frac{d}{(1+k(1-\gamma^{d/k})(1+\lambda+\frac{1}{\lambda}))}\right)^2. \end{aligned}$$

The transition rate from the probing state to the transmitting state is $k(1-\gamma^{d/k})$. Note that $k(1-\gamma^{d/k})$ is increasing in k when $\frac{d}{k} \geq 1$ because

$$\frac{\partial}{\partial k}\left(k\left(1-\gamma^{\frac{d}{k}}\right)\right) = 1 - \gamma^{\frac{d}{k}} + \gamma^{\frac{d}{k}}\frac{d}{k}\log\gamma.$$

Now define

$$f(y, \gamma) = 1 - \gamma^y + \gamma^y y \log \gamma.$$

We next prove that $f(y) > 0$ for $y \geq 1$ and $0 < \gamma \leq 1$. Note that

$$\frac{\partial}{\partial y} f(y, \gamma) = -\gamma^y \log \gamma + \gamma^y \log \gamma + \gamma^y y (\log \gamma)^2 = \gamma^y y (\log \gamma)^2 > 0.$$

Now consider

$$f(1, \gamma) = 1 - \gamma + \gamma \log \gamma.$$

We have

$$\frac{\partial}{\partial \gamma} f(1, \gamma) = \log \gamma < 0.$$

Therefore, we conclude that for $y \geq 1$ and $0 < \gamma \leq 1$, we have

$$f(y, \gamma) > f(1, \gamma) \geq f(1, 1) = 0,$$

i.e.

$$\frac{\partial}{\partial k} \left(k \left(1 - \gamma^{\frac{d}{k}} \right) \right) = 1 - \gamma^{\frac{d}{k}} + \gamma^{\frac{d}{k}} \frac{d}{k} \log \gamma > 0$$

Define $x = k(1 - \gamma^{d/k})$. We obtain

$$J(x) = -\frac{(1+\lambda)}{\frac{1}{x} + 1 + \lambda + \frac{1}{\lambda}} + c \left(\frac{d}{1 + x(1 + \lambda + \frac{1}{\lambda})} \right)^2,$$

which is clearly a decreasing function of x. Therefore, for fixed d, $J(d, k)$ is a decreasing function of k. Therefore, we have $J(d, d) < J(d, k)$ when $d > k$. $\square$

According to the lemma above, given γ, the optimal policy (d^*, k^*) satisfies $k^* = d^*$. In other words, given d, it is optimal to probe one channel at a time with rate d. Therefore, in the following discussion, we focus on probing policies such that $d = k$.

Since $d = k$, we will now proceed assuming that each device wishes to optimize a cost function written in terms of d. This function can be written as:

$$\begin{aligned} J(d) = &- \frac{(1+\lambda)d(1-\gamma)}{1+d(1-\gamma)(1+\lambda+\frac{1}{\lambda})} \\ &+ c\left(\frac{d}{1+d(1-\gamma)(1+\lambda+\frac{1}{\lambda})}\right)^2. \end{aligned} \tag{2.7}$$

and the dynamical system can be written as:

$$\begin{aligned} \frac{dq_0}{dt} &= -\lambda q_0 + \frac{1}{1+\lambda} q_2 \\ \frac{dq_1}{dt} &= \lambda q_0 - d(1-mq_2)q_1 \\ \frac{dq_2}{dt} &= d(1-mq_2)q_1 - \frac{1}{1+\lambda} q_2 \end{aligned} \tag{2.8}$$

Theorem 1. *Assume that all devices use the same policy* (d, d)*. Let* $\gamma^{(N)}(\infty)$ *denote the fraction of busy channels at the steady state in a system with* N *channels and* mN *devices. Then* $\gamma^{(N)}(\infty)$ *converges weakly to* γ*, which is the unique equilibrium of mean-field model* (2.4) *with* $d = k$*, and is the unique solution of the following equation:*

$$\gamma = \frac{m(1+\lambda)k(1-\gamma)}{1+d(1-\gamma)(1+\lambda+\frac{1}{\lambda})}. \tag{2.9}$$

Due to the lack of space we restrict the proof of convergence to the appendix, where we also briefly discuss the derivation of the mean-field model (2.4). Figure 2.3 shows the simulation results with $m = 5$, and $c = 10$, $\lambda = 0.7$, and $d = 0.065$. We varied N from 10, to 100 and then to 1,000. We can clearly see that γ converges to the mean-field limit as N increases, and when $N = 1,000$, γ concentrates to the mean-field limit.

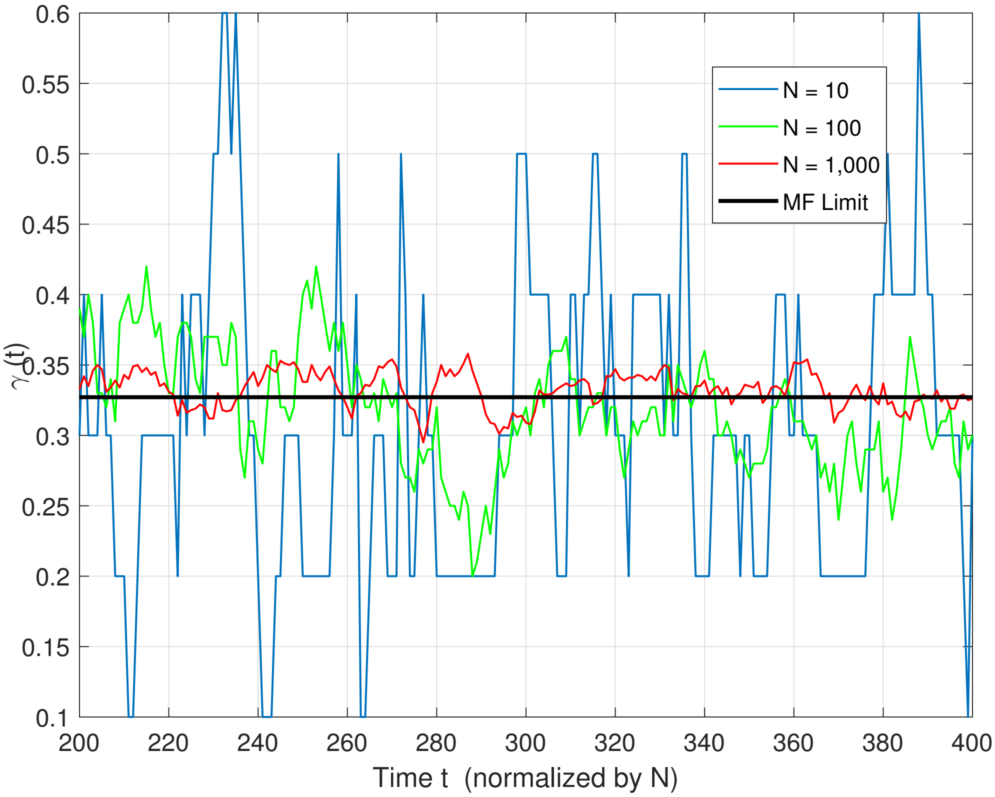


Figure 2.3: Convergence to the Mean Field Limit with Fixed d

2.4 Uniqueness and Convergence of MFNE

In the previous section, we have shown that given policy (d, d), the stationary distribution of the mN-device system converges to a unique mean-field limit, which defines mapping

$$T_1 : d \to \gamma. \tag{2.10}$$

The mapping

$$T_2 : \gamma \to d \tag{2.11}$$

is obtained by solving the optimization problem $\min_k J(d)$ for given γ.

The following lemma provides the closed-form expression of mapping T_2.

Lemma 2. *Given* $0 < \gamma < 1$ *and* $d \geq 0$, $J(d)$ *has a unique minimizer*

$$d = \frac{a}{\max\{2c - ab, 0\}},$$

where $a = (1-\gamma)(1+\lambda)$ *and* $b = (1-\gamma)\left(1 + \lambda + \frac{1}{\lambda}\right)$.

Proof. Define $a = (1+\lambda)(1-\gamma)$ and $b = (1-\gamma)(1+\lambda+\frac{1}{\lambda})$. Then $J(d)$ can be written as

$$J(d) = -\frac{ad}{1+bd} + c\left(\frac{d}{1+bd}\right)^2,$$

and

$$\frac{\partial J(d)}{\partial d} == \frac{1}{(1+bd)^2}\left(-a + \frac{2cd}{1+bd}\right).$$

We now consider

$$h(d) = -a + \frac{2cd}{1+bd}.$$

Note that $h(d)$ is an increasing function for $d \geq 0$. Furthermore $h(0) = -a$ and

$$h(d) \leq \lim_{d\to\infty} h(d) = -a + \frac{2c}{b}.$$

Therefore, if $\frac{2c}{b} \leq a$, (i.e. $h(d) \leq 0$), then $J(d)$ is a strictly decreasing function and the minimum is achieved at $d = \infty$. Otherwise, the minimum is achieved when

$$d = \frac{a}{2c - ab}.$$

In summary, $J(d)$ is minimized at

$$d = \frac{a}{\max\{2c - ab, 0\}}.$$

□

Now given mapping T_1 characterized in Theorem 1 and mapping T_2 characterized in Lemma 2, the following theorem establishes the existence and uniqueness of the MFNE.

Theorem 2. *The existence of MFG equilibria depends on the traffic load* λ *and constant* c. *The results can be divided into three cases. For fixed* c, *the following three cases correspond to "low", "high" and "medium" traffic regimes.*

- **Case I (Low Traffic Regime)**: *If*

$$2c \leq \left(\max\left\{0, 1 - \frac{m(1+\lambda)}{1+\lambda+\frac{1}{\lambda}}\right\}\right)^2 (1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right), \qquad (2.12)$$

  then $d^* = \infty$ *is the unique MGF equilibrium. In other words, in this case, a device should continuously probe idle channels (with no waiting) when there is a message to transmit.*

- **Case II (High Traffic Regime)**: *If*

$$2c > (1-\gamma^*)^2 (1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right), \qquad (2.13)$$

  where

$$\gamma^* = 1 + \frac{c}{m(1+\lambda)^2} - \sqrt{\frac{c^2}{m^2(1+\lambda)^4} + \frac{2c}{m(1+\lambda)^2}},$$

  then there exists a unique MGF equilibrium

$$d^* = \frac{(1-\gamma^*)(1+\lambda)}{2c - (1-\gamma^*)^2(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)}. \qquad (2.14)$$

- **Case III (Medium Traffic Regime)**: *Otherwise, MFNE does not exist and devices switch probing strategy between* $d = \infty$ *and*

$$d = \frac{(1-\tilde{\gamma})(1+\lambda)}{2c - (1-\tilde{\gamma})^2(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)},$$

  where

$$\tilde{\gamma} = \min\left\{1, \frac{m(1+\lambda)}{1+\lambda+\frac{1}{\lambda}}\right\}.$$

Proof. We first consider Case I such that

$$2c \leq \left(\max\left\{0, 1 - \frac{m(1+\lambda)}{1+\lambda+\frac{1}{\lambda}}\right\}\right)^2 (1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right). \tag{2.15}$$

Under this condition, we have

$$1 - \frac{m(1+\lambda)}{1+\lambda+\frac{1}{\lambda}} > 0. \tag{2.16}$$

Recall (q_0^*, q_1^*, q_2^*) denote the unique equilibrium point of mean field model (A.5) for a given d. For any $d \geq 0$, we have

$$q_2^* \leq \frac{1+\lambda}{1+\lambda+\frac{1}{\lambda}}.$$

This upper bound holds because the following equations holds for all $d > 0$:

$$\lambda q_0^* = \frac{1}{1+\lambda} q_2^* \tag{2.17}$$

$$\sum_i q_i^* = 1, \tag{2.18}$$

which implies

$$\frac{\frac{1}{\lambda}}{1+\lambda} q_2^* + q_1^* + q_2^* = 1$$

and

$$\left(1 + \frac{\frac{1}{\lambda}}{1+\lambda}\right) q_2^* \leq 1.$$

Recall that $\gamma^* = m q_2^*$, so

$$\gamma^* \leq \frac{m(1+\lambda)}{1+\lambda+\frac{1}{\lambda}}.$$

Substituting this inequality into (2.15), we have that the following inequality holds for any $d \geq 0$:

$$2c \leq (1-\gamma^*)^2 (1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right) = ab, \tag{2.19}$$

where a and b are defined in Lemma 2. Therefore, $2c \leq ab$, and $d^* = \infty$ according to Lemma 2. Furthermore, given $d^* = \infty$, we have

$$\gamma^* = \frac{m(1+\lambda)}{1+\lambda+\frac{1}{\lambda}} > 0$$

according to Theorem 1 by taking $d \to \infty$. Therefore, $d^* = \infty$ is the unique MFG equilibrium.

Now if $d^* < \infty$ is a MFG equilibrium, it satisfies the following two equations

$$\begin{aligned} d^* =& \frac{(1-\gamma^*)(1+\lambda)}{2c - (1-\gamma^*)^2(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)} \\ \gamma^* =& \frac{md^*(1-\gamma^*)(1+\lambda)}{1+d^*(1-\gamma^*)(1+\lambda+\frac{1}{\lambda})}. \end{aligned} \tag{2.20}$$

Substituting the first equation into the second one, we obtain

$$\begin{aligned} \gamma^* =& \frac{m(1-\gamma^*)(1+\lambda)\frac{(1-\gamma^*)(1+\lambda)}{2c-(1-\gamma^*)^2(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)}}{1+(1-\gamma^*)(1+\lambda+\frac{1}{\lambda})\frac{(1-\gamma^*)(1+\lambda)}{2c-(1-\gamma^*)^2(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)}} \\ =& \frac{m(1+\lambda)^2}{2c}(1-\gamma^*)^2. \end{aligned}$$

Note that $\gamma^* = \frac{m(1+\lambda)^2}{2c}(1-\gamma^*)^2$ has a unique solution $\gamma^* \in (0,1)$ since γ^* is an increasing function (increasing from 0 to 1) and $(1-\gamma^*)^2$ is a decreasing function (decreasing from 1 to 0). In particular, the unique solution is

$$\gamma^* = 1 + \frac{c}{m(1+\lambda)^2} - \sqrt{\frac{c^2}{m^2(1+\lambda)^4} + \frac{2c}{m(1+\lambda)^2}}. \tag{2.21}$$

Now to guarantee $d^* < \infty$, it requires

$$2c > (1-\gamma^*)^2(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)$$

according to (2.20), which concludes Case II.

Finally we consider Case III. When condition

$$2c > (1-\gamma^*)^2(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)$$

does not hold, after learning γ^* defined in (2.21), all devices choose strategy $d = \infty$. However, when

$$2c > \left(\max\left\{0, 1 - \frac{m(1+\lambda)}{1+\lambda+\frac{1}{\lambda}}\right\}\right)^2 (1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right), \tag{2.22}$$

$d = \infty$ is not an MFG equilibrium because

$$\tilde{\gamma} = T_1(\infty) = \min\left\{1, \frac{m(1+\lambda)}{1+\lambda+\frac{1}{\lambda}}\right\}$$

but

$$\tilde{d} = T_2\left(\tilde{\gamma}\right) < \infty$$

when

$$2c > ab = (1-\tilde{\gamma})^2(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right).$$

Therefore, after all devices choosing $d = \infty$, the fraction of busy channels is $\tilde{\gamma}$ in the mean-field limit. After learning the fraction of busy channels is $\tilde{\gamma}$, all devices change their policy to $d = \tilde{d}$. It can be verified that $T_1(\tilde{d}) \leq \gamma^*$, so under policy $\tilde{d}$, the fraction of busy channels in the mean-field limit is at most γ^*. Then after learning the fraction of busy channels, all devices switch to policy $d = \infty$. Therefore no MFG equilibrium exists in this case. The system switches between $d = \infty$ and $d = \tilde{d}$. $\square$

The theorem above presents the conditions under which an MFNE exits. Next, we study the convergence (i.e, stability) of the MFNE. For Case I, the convergence is immediate as indicated in the proof of Theorem 2, where we can see that all devices choose strategy $d^* = \infty$ after learning the fraction of busy channels and reach the MFNE. We now focus on Case II under which d^* is a finite value and have the following global convergence result. Since no MFNE exits in Case III, the question of convergence is irrelevant.

Theorem 3. *Consider Case II in Theorem 2. For any $c > c_{m,\lambda}$ where $c_{m,\lambda}$ is a positive constant such that*

$$m\frac{(1+\lambda)}{(1+\lambda+1/\lambda)}\frac{\frac{2c_{m,\lambda}}{(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)}+1}{\left(\frac{2c_{m,\lambda}}{(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)}-1\right)^2} = 1,$$

the system converges to the MFNE starting from any initial condition.

We remark that convergence to the mean-field limit (Theorem 1) and convergence to the MFNE (Theorem 3) are two fundamentally different concepts. Convergence to the mean-field limit shows that the stationary distributions of finite size systems converge weakly to the fixed point of the mean-field model for *fixed* (d, k), so no "game" is involved but the result does justify the MFG approach. On contrast, convergence to the MFNE does not involve finite-size stochastic systems, but investigates the dynamics of the MFG. The result shows that the iterative process, defined as the MFG for distributed MAC in Section 2.2, converges to the unique MFNE.

Proof. Recall mappings T_1 and T_2. Given policy (d, d), the stationary distribution of the mN-device system converges to a unique mean-field limit, which defines the following mapping

$$T_1 : d \to \gamma. \tag{2.23}$$

The mapping

$$T_2 : \gamma \to d \tag{2.24}$$

is obtained by solving the optimization problem $\min_d J(d)$ for given γ.

We begin by showing that, for fixed m, T_1 always has Lipschitz constant which is upper bounded by $m(1 + \lambda)$. Based on (2.9), we first obtain

$$\begin{aligned}\frac{\partial\gamma}{\partial d} = &- \frac{m(1+\lambda)k}{\big(1 + d(1-\gamma)(1+\lambda+\frac{1}{\lambda})\big)^2}\frac{\partial\gamma}{\partial d} \\ &+ \frac{m(1+\lambda)(1-\gamma)}{\big(1 + k(1-\gamma)(1+\lambda+\frac{1}{\lambda})\big)^2}\end{aligned}$$

which implies that

$$\begin{aligned}\left|\frac{\partial\gamma}{\partial d}\right| &= \frac{m(1+\lambda)(1-\gamma)}{m(1+\lambda)d + \big(1 + d(1-\gamma)(1+\lambda+1/\lambda)\big)^2} \\ &< m(1+\lambda)(1-\gamma) \\ &< m(1+\lambda).\end{aligned}$$

Recall that T_2 is a map from γ to d which gives us the unique minimizer for the cost function $J(d)$, and that we consider Case II such that

$$2c > (1-\gamma)^2(1+\lambda)\left(1+\lambda+\frac{1}{1+\lambda}\right),$$

and

$$d = \frac{(1-\gamma)(1+\lambda)}{2c-(1-\gamma)^2(1+\lambda)(1+\lambda+1/\lambda)}.$$

Define $\alpha = \frac{2c}{(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)}$, we further obtain

$$k = \frac{1}{(1+\lambda+1/\lambda)}\frac{1-\gamma}{\alpha-(1-\gamma)^2},$$

from which, we have

$$\begin{aligned}\left|\frac{\partial d}{\partial \gamma}\right| =& \frac{1}{1+\lambda+1/\lambda}\frac{\alpha+(1-\gamma)^2}{(\alpha-(1-\gamma)^2)^2}\\ <& \frac{1}{1+\lambda+1/\lambda}\frac{\alpha+1}{(\alpha-1)^2}.\end{aligned}$$

Define $T(d) = T_2(T_1(d))$. From the discussion above, we have

$$\frac{\partial T}{\partial d} = \left|\frac{\partial d}{\partial \gamma}\right|\left|\frac{\partial \gamma}{\partial d}\right| \leq m\frac{(1+\lambda)}{(1+\lambda+1/\lambda)}\frac{\alpha+1}{(\alpha-1)^2}.$$

Note

$$\frac{\alpha+1}{(\alpha-1)^2}$$

is a decreasing function of α for $\alpha > 1$ because

$$\frac{d}{d\alpha}\left(\frac{\alpha+1}{(\alpha-1)^2}\right) = -\frac{\alpha+3}{(\alpha-1)^3} < 0,$$

so is a decreasing function of c according to the definition of α. Furthermore,

$$\lim_{\alpha\to\infty}\frac{\alpha+1}{(\alpha-1)^2} = 0.$$

Therefore, given m and λ, there exists $c_{m,\lambda} > 0$ such that

$$m\frac{(1+\lambda)}{(1+\lambda+1/\lambda)}\frac{\frac{2c_{m,\lambda}}{(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)}+1}{\left(\frac{2c_{m,\lambda}}{(1+\lambda)\left(1+\lambda+\frac{1}{\lambda}\right)}-1\right)^2} = 1.$$

For any $c > c_{m,\lambda}$, we have a contraction mapping and the system converges to the MFG equilibrium. □

The following theorem further shows that under the conditions of Theorem 3, the MFNE is an ϵ-Nash-equilibrium such that if all other players use sampling rate at the MFNE, then the cost of a player who uses a different sampling rate can deviate no more than $O(\epsilon)$ from the cost of using the sampling rate at the MFNE.

While we did show that under the conditions of Theorem 3, a unique fixed point exists, this does not necessarily imply that a given user will follow the MFNE strategy. We need to show that the fixed point is a Nash equilibrium, this means that a given user cannot hope to benefit through unilateral deviation. Our proof technique shows that if a single user chooses to deviate, the fraction of busy channels will deviate by a factor that goes to zero as M tends to infinity. Note, this is a nontrivial proof because we have to deal with a heterogeneous system without any asymptotic independence condition. Next given the fraction of busy channels, the probing rate is chosen optimally for utility J, and since J is Lipschitz with respect to γ, this means that the new probing rate will also deviate by a factor that goes to zero as M tends to infinity. The result is proved using the result on the approximation error of mean-field models Ying (2016).

Theorem 4. *Under the conditions of Theorem 3, the $MFNE$ is an ϵ-Nash-Equilibrium of the M-player Baysian game, where $\epsilon = O\left(\frac{1}{M^{1/3}}\right)$.*

Proof. Under the conditions of Theorem 3, we consider two systems.

- System 1: In this system, all players use sampling rate d^* defined in (2.14). Let $\gamma^{(N)}(t)$ denote the fraction of busy channels of the system at time t, which is a random variable, and $\gamma^{(N)}(\infty)$ denote the fraction of busy channels at steady state.

- System 2: The first $M-1$ players use sampling rate d^* and the Mth player uses a different sampling rate. Let $\tilde{\gamma}^{(N)}(t)$ denote the fraction of busy channels of the system at time t, which is a random variable, and $\tilde{\gamma}^{(N)}(\infty)$ denote the fraction of busy channels at steady state.

Now to prove the theorem, we first show that the fraction of busy channels $\tilde{\gamma}^{(N)}(\infty)$ remains to be close to γ^* even when the Mth player uses a sampling rate different from d^*.

Lemma 3.

$$E\left[\left\|\tilde{\gamma}^{(N)}(\infty)-\gamma^*\right\|^2\right] \leq \frac{C}{M} \tag{2.25}$$

for some positive constant C independent of M. □

The proof of the lemma can be found in the Appendix.

From lemma 3 and Chebychev's inequality, we can obtain

$$\Pr(\|\gamma^*-\tilde{\gamma}^{(N)}(\infty)\| > \epsilon) \leq \frac{C^2}{\epsilon^2 M}.$$

We now compare the costs of the Mth player in system 1 (using probing rate $d^* = T_2(\gamma^*)$) and in system 2 (using a different sampling rate α). The policy picked by any device given fraction of busy channels $mq_2(\infty)$ is $T_2(mq_2(\infty))$. The policy picked by the $MFNE$ policy is $T_2(\gamma*)$.

Given $\gamma > 0$, the cost function $J(d)$ is Lipschitz in d (see the proof of Lemma 2). Therefore, for any γ, the map

$$\gamma \xrightarrow{T_2} d \rightarrow J(d).$$

is also Lipschitz in γ. Say, the Lipschitz constant is L, then

$$\begin{aligned}
&E\left[J(T_2(\gamma^*)) - J(T_2(\tilde{\gamma}^{(N)}(\infty)))\right] \\
\leq& LE\left[\gamma^* - \tilde{\gamma}^{(N)}(\infty))\right] \\
\leq& LE\left[\gamma^* - \tilde{\gamma}^{(N)}(\infty)\middle|\, \|\gamma^* - \tilde{\gamma}^{(N)}(\infty)\| > \epsilon\right] \times \\
& \ \Pr\left(\|\gamma^* - \tilde{\gamma}^{(N)}(\infty)\| > \epsilon\right) + \\
& \ LE\left[\gamma^* - \tilde{\gamma}^{(N)}(\infty)\middle|\, \|\gamma^* - \tilde{\gamma}^{(N)}(\infty)\| \leq \epsilon\right] \\
\leq& \frac{2LC^2}{\epsilon^2 M} + L\epsilon.
\end{aligned}$$

Choosing $\epsilon = \frac{1}{M^{1/3}}$,

$$E\left[J(T_2(\gamma^*)) - J(T_2(\tilde{\gamma}^{(N)}(\infty)))\right] = O\left(\frac{1}{M^{1/3}}\right).$$

This concludes the proof.

□

2.5 Price of Anarchy

In this section, we analyse the performance of the distributed MAC with respect to a global optimal solution where a centralized controller chooses the optimal k for minimizing

$$\begin{aligned}
\hat{J}(d) =& -\frac{(1+\lambda)d(1-\gamma)}{1+d(1-\gamma)(1+\lambda+\frac{1}{\lambda})} + \\
& c\left(\frac{d}{1+d(1-\gamma)(1+\lambda+\frac{1}{\lambda})}\right)^2,
\end{aligned} \tag{2.26}$$

where

$$\gamma = \frac{m(1+\lambda)d(1-\gamma)}{1+d(1-\gamma)(1+\lambda+\frac{1}{\lambda})}. \tag{2.27}$$

Denote by $\hat{d}$ the optimal solution. All devices are forced to use probing rate $\hat{d}$. We will call the cost corresponding to this probing rate the global optimal cost and compare it with the cost at the MFNE.

Recall that for the MFNE, each device minimizes it cost function by assuming that γ is fixed. For the centralized case, the controller solves (2.5) by considering γ to be a function of d as defined in (2.27). This is the reason the global optimal solution differs from the cost at the MFNE. Let $\hat{\gamma}$ denote the fraction of busy channels that occurs as a result of the central controller picking an optimal sampling rate. Define

$$1 - \frac{|J(\gamma^*)|}{|\hat{J}(\hat{\gamma})|}$$

to be the price of anarchy. The following theorem shows that the price of anarchy is at most 0.5. The price of anarchy can be viewed as a measure of efficiency when comparing a distributed protocol to a centralized protocol. It measures the loss of utility that occurs when we pick a distributed implementation instead of a central one. Note that the cost at the MFNE and the global optimal cost are both negative because the policy that does not probe any channel and does not transmit any message has cost zero. Therefore, lower the cost, the larger its absolute value. In the following proof, we characterize the price-of-anarchy by analyzing the conditions $\hat{\gamma}$ and γ^* have to satisfy, which yield

$$\frac{|J(\gamma^*)|}{|\hat{J}(\hat{\gamma})|} = \frac{1}{(1+\hat{\gamma})}\frac{\gamma^*}{\hat{\gamma}}, \tag{2.28}$$

and $\hat{\gamma} < \gamma^*$. Then we can conclude that the price of anarchy is at most 0.5.

Theorem 5. *The price of anarchy,* $1 - |J(\gamma^*)|/|\hat{J}(\hat{\gamma})|$, *is at most* $1/2$. *In Case I, the low traffic regime defined in Theorem 2, the price of anarchy is zero.*

We note that in the low traffic regime (Case I in Theorem 2), both the distribution MAC and the centralized solution use probing strategy with $k^* = \infty$, so the price of anarchy is zero. We provide a proof for Case II defined in Theorem 2.

Proof. By substituting (2.27) into (2.5), we obtain

$$\hat{J}(\gamma) = -\frac{\gamma}{m} + c\left(\frac{\gamma}{m(1+\lambda)(1-\gamma)}\right)^2$$

The optimal solution to minimize $\hat{J}(\gamma)$ can be obtained by setting $\frac{\partial \hat{J}}{\partial \gamma}$ to be zero, which yields that the minimizer $\hat{\gamma}$ is the unique solution to the following equation

$$\hat{\gamma} = \frac{m(1+\lambda)^2(1-\hat{\gamma})^3}{2c}.$$

By simple substitution, we further obtain

$$\hat{J}(\hat{\gamma}) = -\frac{(1+\lambda)^2}{4c}(1-\hat{\gamma})^3(1+\hat{\gamma}) \tag{2.29}$$

It can be shown (and indeed we show this in the appendix) that γ^* is the unique solution of the following equation

$$\gamma^* = \frac{m(1+\lambda)^2(1-\gamma^*)^2}{2c}.$$

By substituting it into (2.5), we have

$$J(\gamma^*) = -\frac{(1+\lambda)^2}{4c}(1-\gamma^*)^2. \tag{2.30}$$

The ratio of the cost function at MFNE to the optimal cost function is given by:

$$\frac{|J(\gamma^*)|}{|\hat{J}(\hat{\gamma})|} = \frac{1}{(1+\hat{\gamma})}\frac{(1-\gamma^*)^2}{(1-\hat{\gamma})^3} = \frac{1}{(1+\hat{\gamma})}\frac{\gamma^*}{\hat{\gamma}}, \tag{2.31}$$

where the last equality holds because

$$\frac{\gamma^*}{\hat{\gamma}} = \frac{\frac{m(1+\lambda)^2(1-\gamma^*)^2}{2c}}{\frac{m(1+\lambda)^2(1-\hat{\gamma})^3}{2c}} = \frac{(1-\gamma^*)^2}{(1-\hat{\gamma})^3}.$$

Observe that $\hat{\gamma}$ is strictly smaller than γ^* because otherwise

$$\frac{\gamma^*}{\hat{\gamma}} < \frac{(1-\gamma^*)^2}{(1-\hat{\gamma})^3}.$$

Therefore, we conclude that

$$1 > \frac{|J(\gamma^*)|}{|\hat{J}(\hat{\gamma})|} > \frac{1}{(1+\hat{\gamma})} > \frac{1}{2}.$$

Which implies that:

$$0 < \text{Price of Anarchy} < \frac{1}{2}$$

In other words, the price of anarchy is upper bounded by 0.5. $\square$

Focusing on Case II defined in Theorem 2, Figure 2.4 shows the price of anarchy with $c = 0.1$ and $m = 5$ with λ varying from 0.5 to 2. We can see that the price of anarchy increases as λ increases and approaches 0.5.

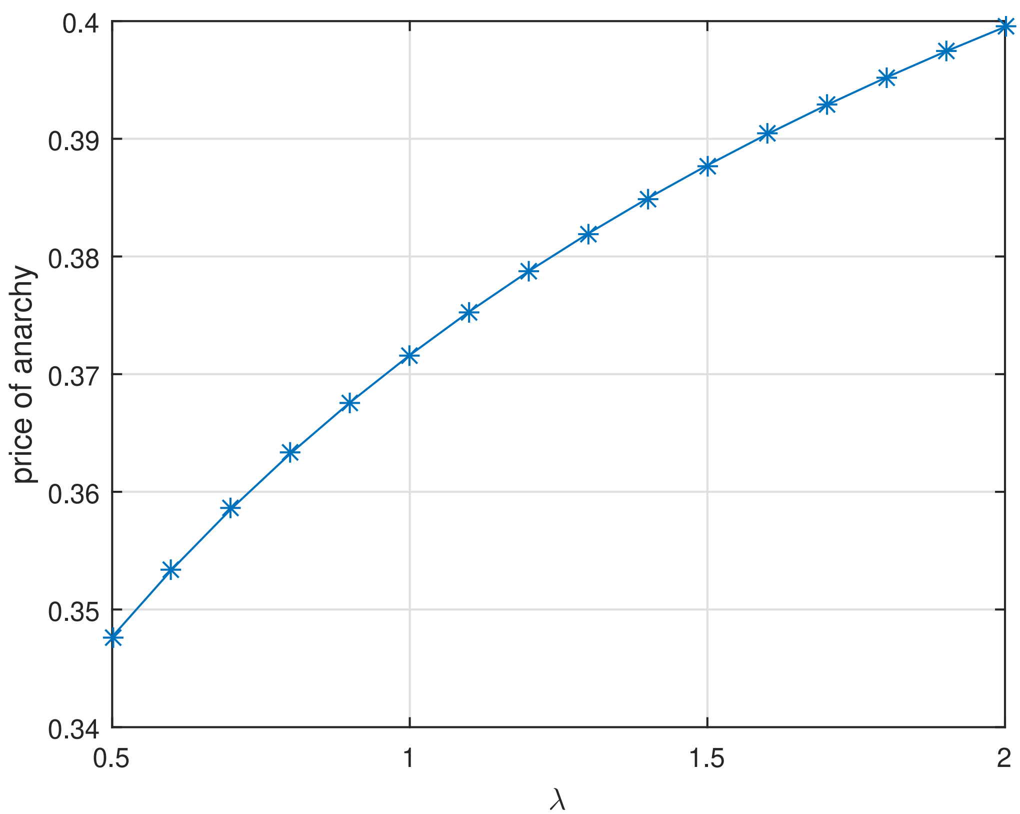


Figure 2.4: Price of Anarchy versus λ

2.6 Simulations

In this section, we use simulations to compare the distributed MAC policy, named DMAC-G for short with other similar light-weight distributed protocols. We simulated $N = 1,000$ devices with $m = 5$, and $c = 10$, and the average λ varying from 0.5 to 1. These choices of parameters guarantee the existence and convergence to the MFNE. We used uniformization to simulate the CTMC described in our system model in Section 2.1.

DMAC -G : We simulated two different scenarios for the DMAC -G protocol:homogeneous case where all devices have the same arrival rate and the same parameter, c and heterogeneous case where devices have different arrival rates and different values of parameter c. Since we ran the simulations on a laptop without parallelization, to speed up the simulations, the fraction of busy channels was measured as a common variable shared by all devices. In this way, we were able to simulate an M-device system efficiently using uniformization.

- **The homogeneous case** In the homogeneous case every device has the same arrival rate λ and energy parameter c. Hence, each device has the same utility function and so will choose the same sampling rate when given the common random variable for the fraction of busy channels.

- **The heterogeneous case** Each device follows the policy (d, d), however, the devices have different arrival rates and parameters c. The arrival rates were picked uniformly at random from $[0.75\lambda, 1.25\lambda]$. Similarly the values of the parameter c were chosen uniformly at random from $[0.75c, 1.25c]$ for some c.

Therefore, both cases have the same average arrival rates and cost parameters. Figures 2.5, 2.6 and 2.7 show that both scenarios yield very similar cost, fraction of busy channels and delay. We compare our algorithm under these two scenarios with a CSMA protocol with exponential back off.

E-CSMA : Each device maintains an exponential clock with initial rate $k = 1$. When the clock ticks, the device probes one of the N channels, chosen uniformly at random. If the probed channel is idle, the device starts to transmit the packet, if not the device halves its sampling rate and the clock restarts. We simulated this protocol under both homogeneous and heterogeneous scenarios.

We evaluated the performance of the protocols in terms of the cost and per-packet delay(for those successfully transmitted packets). We can observe from Figure 2.5 that DMAC-G yield a lower cost than E-CSMA and the gap increases as λ increases. Note, that the cost function is a linear combination of the probing cost minus the throughput. From Figure 2.7, we can also observe that our algorithm has much lower per-packet delay. The average delay is less than 2 for all the λ under DMAC-G, which reduces the probing rate when the traffic load increases, which reduces overall cost and per-packet delay(increases the freshness of the information).

These simulations confirm: (i) the analytical results in this report, while derived for the homogeneous case, also match the performances of the heterogeneous case reasonably well; and (ii) our low-complexity, adaptive MAC protocol signficantly outperforms the exponential back-off MAC protocol (a commonly used MAC protocol).

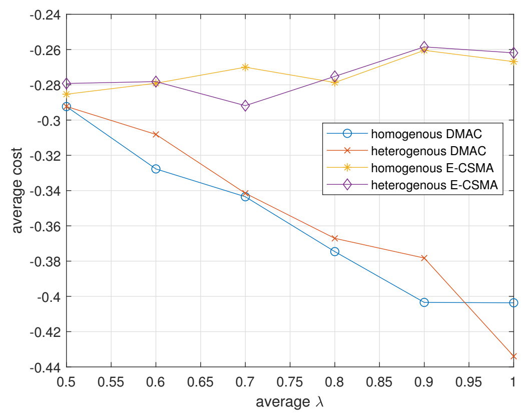


Figure 2.5: Average Costs under the Four Different Scenarios.

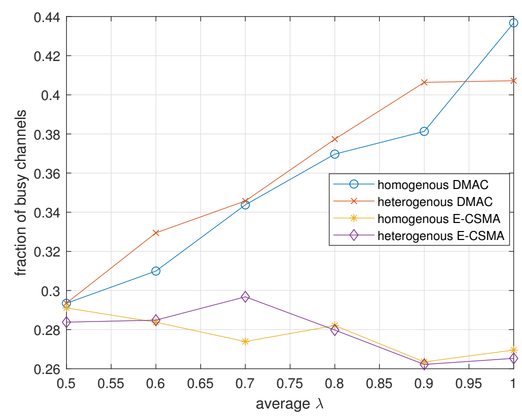


Figure 2.6: Average Fraction of Busy Channels under the Four Different Scenarios.

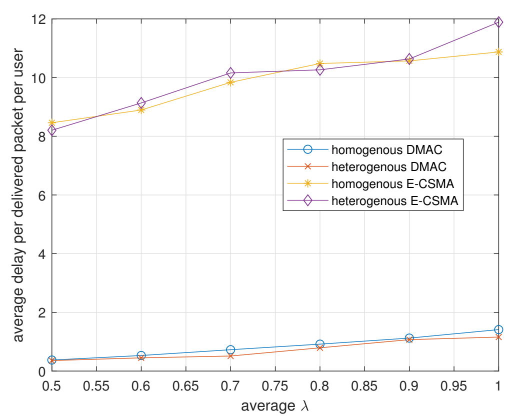


Figure 2.7: Average delay per delivered packet per user under the Four Different Scenarios.

Chapter 3

AGE DEPENDENT MAC

3.1 System Model and Mean-Field Game

We consider an N-channel, M-device ultra dense wireless network. We consider the case when both M and N tend to infinity and M/N is a constant. Each device in the network generates status messages following a Poisson process with rate λ. As we mentioned in the introduction, the age of a message evolves following an exponential clock which ticks with rate $1/\delta$. We call this clock the delay clock. When the delay clock ticks, the age/delay of the message increases from i to $i+1$. When the device has no message to transmit, we define its state to be -1. When a new message arrives, the device goes to state 0. Therefore, the state of the device changes from i to state $i+1$ with rate $1/\delta$.

Each device maintains a separate exponential clock with rate $\alpha_i \in (0, A)$ for some constant A when in state i. When this clock ticks, the device will probe one channel at random to see if it is free. If the channel is free, the device grabs the channel and starts transmitting its message. In this case, the state of the device moves from i to state -2 which indicates that it is in the transmitting state.If a new message arrives when the device is in the probing state, the age of message is reset to 0, which means that the state of the device transits from state i to state 0. On the other hand if a new message arrives while the device is transmitting its message, the message is stored and transmitted immediately after the current message without giving up the channel. The state space diagram for this system in Fig 3.1.

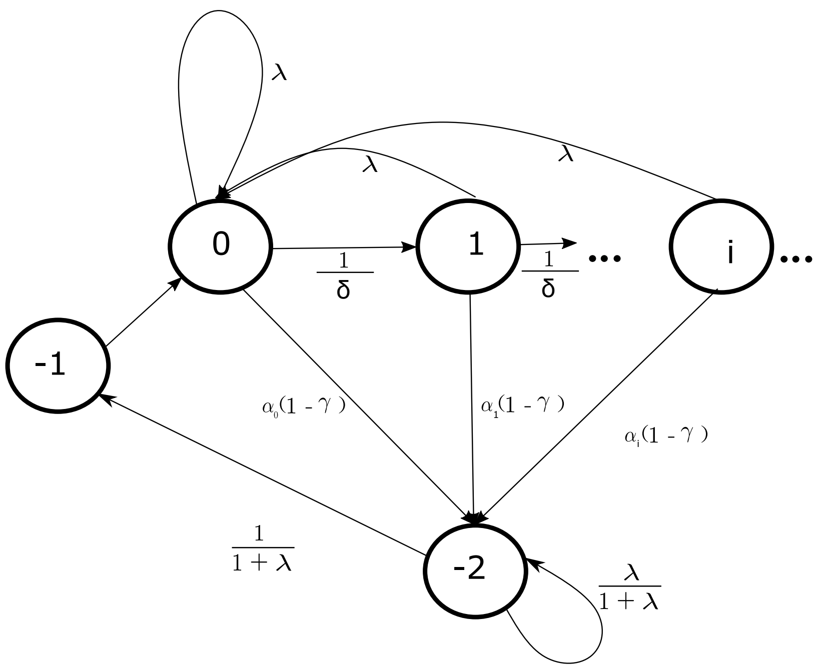


Figure 3.1: State space model for the Markov chain

A device receives a reward of R_i for transmitting a message with age i (i.e. at state i). Like in most IoT or status messaging applications, it is important for devices to transmit their messages with fresh information, so the reward R_i decreases in i and decreases to 0 as i increases. Further, we assume that for messages that arrive while the device is in state -2, the reward for transmitting each of these messages is a constant r_{-2}. Additionally, with probing rate α, the device needs to pay a cost of $\hat{c}(\alpha)$, which is a strictly increasing function with $\hat{c}(0) = 0$. Consider the corresponding jump process for the CTMC described above, then we use t_j to denote the time between the j^{th} tick and the $j + 1^{st}$ tick of the overall exponential clock. Note that with probing rate α, the expected transition time is $\frac{1}{\alpha+\lambda+\frac{1}{\delta}}$.

Given the model presented above, each device maximizes the following discounted infinite-horizon problem:

$$
\begin{aligned}
u_x =& \frac{1}{A+\lambda+1/\delta} \times \\
& \max_{\alpha\in[0,A]} \left(\alpha+\lambda+\frac{1}{\delta}\right) E\left[\sum_{j=0}^{\infty} \beta^j \left(1_{X(j+1)=-2} R_{X(j)} \right.\right. \\
& \left.\left. - 1_{\{X(j+1)=X(j) \bigcup X(j+1)=-2\}} \hat{c}(\alpha_{X(j)})\right) \middle| X(0)=x \right],
\end{aligned} \tag{3.1}
$$

where β is the discount factor, $X(0)$ is the initial state of the CTMC, and $X(j)$ is the state of the CTMC after the j^{th} transition. We can view this Bellman equation as the normalized time averaged reward that each device obtains when initialized with state x with a constant $(A+\lambda+1/\delta)$. If one imagines each device to maintain a super clock used to simulate all the events, then this super clock will need to have a tick rate of $(A+\lambda+1/\delta)$. One can therefore view $\frac{1}{A+\lambda+1/\delta}$ as a normalized unit of time. The time spent in state $X(j)$ before the next event is given by $\frac{1}{\alpha+\lambda+1/\delta}$ with probing clock ticks with rate α.

Note that, in order for a given device to find the optimal policy it must take into account the fraction of busy channels, $\gamma(t)$. Which itself is a density dependent random process that is determined by the states of all the devices. In other words, this fraction of busy channels $\gamma(t)$ couples all M devices, and makes it intractable to solve the steady-state of the system and the optimal policy based on the Bellman equation.

3.1.1 Mean Field Game

To overcome this difficulty, we approach the problem from a mean-field game perspective. We assume the time-scale separation (a similar assumption used Narasimha *et al.* (2020)) such that the devices adapt their policies in a slower time-scale than the convergence of the system to its steady-state with a fixed policy.

Under this time-scale separation, with a fixed policy α for all devices, the steady-state of the stochastic system converges to the equilibrium point of a mean field model, called the mean-field-limit, to be defined in Section 3.2 as the system size increases. In particular, the fraction of busy channels $\gamma(t)$ converges to a point mass γ, i.e. a constant. Let us denote this mapping under this mean-field limit when the policy α is given by T_1 such that

$$T_1 : \alpha \to \gamma.$$

This occurs at a *fast* time-scale.

Now, for the fixed γ under the previous policy, each device solves the Bellman equation to determine a new policy for the given γ. Thus, for a fixed γ, each device finds its policy based on the Bellman equation with a constant γ (see details on the structure of the policy in Section 3.3). We denote this mapping by T_2 such that

$$T_2 : \gamma \to \alpha.$$

We now define an Mean-Field-Nash-Equilibrium (MFNE) as a policy, α^*, such that

$$\alpha^* = T_2(T_1(\alpha^*)),$$

i.e. a fixed point of mapping $T_2(T_1(\cdot))$. In order to show the existence of such a fixed point, we will need to show that the composition of maps T_1 and T_2 is a continuous function. We have already characterized T_1, in the following sections, we will characterize T_2 to show that both T_1 and T_2 are continuous. Hence, the composition of the two maps is also continuous. The existence of a fixed point follows from Brouwer's fixed point theorem. Brouwer's fixed point theorem requires the map from some set $\Omega \to \Omega$ to be continuous and the set Ω must be closed and compact. The last condition can be checked easily, we will prove continuity in Section 3.4 after proving the convergence to the mean-field limit in Section 3.2 and analyzing the policy structure in Section 3.3.

3.2 Mean Field Limit Under a Given Policy: The mapping T_1

This section focuses on the convergence of the stochastic system to its mean-field limit when the policy $\alpha := \{\alpha_0, \alpha_1 ...\}$ is chosen by each device is fixed and such that there exists a finite K such that $\alpha_k = 0$ for all $k > K$. Let $\mathcal{P}^K$ denote the set of all such policies. We will present the necessary assumptions in Section 3.3 so that this condition will be satisfied. Let $Q_j(\infty)$ denote the number of devices in state j at the steady state so $Q_j(\infty)/M$ is the fraction of devices in state j. We further denote S_K to be the fraction of devices who are in state j such that $j \geq K$, i.e. the fraction of devices with delay greater than or equal to K.

In the limit as N and M go to infinity, we will show that the fraction of devices in state $Q_j(\infty)/M$ $j = 1, \cdots, K-1$ and S_K converge weakly to π_j where π_j is the equilibrium point of the mean field model below:

$$\begin{aligned}
\frac{dq_{-1}}{dt} =& -\lambda q_{-1} + \frac{1}{1+\lambda} q_{-2} \\
\frac{dq_{-2}}{dt} =& \sum_{i=0}^{\infty} (1-\gamma)\alpha_i q_i - \frac{1}{1+\lambda} q_{-2} \\
\frac{dq_0}{dt} =& \lambda\big(1 - q_0 - q_{-2}\big) - q_0 \frac{1}{\delta} - \alpha_0(1-\gamma) q_0 \\
\frac{dq_j}{dt} =& \big(q_{j-1} - q_j\big)\frac{1}{\delta} - \alpha_j(1-\gamma)q_j - \lambda q_j \\
& j = 1, \cdots, K-1 \\
\frac{ds_K}{dt} =& q_{K-1}(t)\frac{1}{\delta} - \lambda s_K(t)
\end{aligned} \tag{3.2}$$

Then, if every device fixes this policy α, in the limit as N and M go to infinity we will show that the fraction of devices in state j, $Q_j(\infty)/M$ where Q_j is the number of devices in state j, will converge weakly to π_j where π_j is the equilibrium point of the mean field model (3.3) and $j \in \{-2, -1, 0, 1, 2...\}$. Since, α is in $\mathcal{P}^K$, the mean field dynamical system is given by :

$$\begin{aligned}
\frac{dq_{-1}}{dt} =& -\lambda q_{-1} + \frac{1}{1+\lambda} q_{-2} \\
\frac{dq_{-2}}{dt} =& \sum_{i=0}^{\infty} (1 - mq_{-2}) \alpha_i q_i - \frac{1}{1+\lambda} q_{-2} \\
\frac{dq_0}{dt} =& \lambda \big(1 - q_0 - q_{-2}\big) - q_0 \frac{1}{\delta} - \alpha_0 (1 - mq_{-2}) q_0 . \\
\frac{dq_j}{dt} =& \big(q_{j-1} - q_j\big) \frac{1}{\delta} - \alpha_j (1 - mq_{-2}) q_j - \lambda q_j \\
& j = 1, \cdots, K-1 \\
\frac{dq_j}{dt} =& \big(q_{j-1} - q_j\big) \frac{1}{\delta} - \lambda q_j, \ j \geq K
\end{aligned} \tag{3.3}$$

We can reduce the mean field model above into a finite dimensional dynamical system by truncating the states for delay greater than K.

We will show that for any fixed policy α, the steady-state of the system converges weakly to the solution of the mean field model as $N \to \infty$.

The proof is an application of Theorem 1 in Ying (2016). The theorem states five conditions that are sufficient to guarantee the weak convergence to the fixed point of the mean field model. We next verify these conditions under our model:

- **Bounded transition rate:** This condition can easily be verified from the system model. At any point in time, the rate of transition from any state to any other state is bounded above by $A + \lambda + 1/\delta$.

- **Bounded state transition condition:** Since our model is a collection of M CTMCs whose transition rates are determined by exponential clocks, at most one transition can occur at a time. Therefore, the state transitions are bounded.

- **Perfect Mean Field Model:** Using the system model it can be checked that the equations (3.2) are derived from the detailed balance equations.

- **Partial Derivative condition:** It can be checked that the partial derivatives for the system (3.2) exist and are Lipschitz.

- **Stability conditions:** The global exponential convergence in the appendix in Lemma (9) and Lemma (10).

The following theorem summarizes the result that the Markovian system,

$$\left(\frac{Q_1(\infty)}{M}, \cdots, \frac{Q_{K-1}(\infty)}{M}, S_K(\infty)\right),$$

converges weakly to the fixed point of the dynamical system above,

Theorem 6. *If every device follows a fixed policy α defined in the beginning of this section, then the stationary distribution of the system converges to the unique equilibrium point of system (3.2). This defines the mapping $T_1 : \alpha \to \gamma$.*

Proof. Since, we are showing convergence of the system to the fixed point of dynamical system 3.2, it is standard practice to center the dynamical system. Let π_j for all $j \in \{-2, -1, 0, 1..., K-1\}$and π_{s_K} be the equilibrium point of the dynamical system. Let ϵ_i be the difference between the value of q_i and π_i with $\epsilon_K = s_K - \pi_{s_K}$. The dynamical system written in terms of ϵ_i is given by :

$$\begin{aligned}
\frac{d\epsilon_{-1}}{dt} =& -\lambda\epsilon_{-1} + \frac{1}{1+\lambda}\epsilon_{-2} \\
\frac{d\epsilon_0}{dt} =& -\epsilon_0\big(\frac{1}{\delta} + \alpha_0(1-\gamma) + \lambda\big) + m\alpha_0\pi_0\epsilon_2 - \lambda\epsilon_2 \\
\frac{d\epsilon_j}{dt} =& \frac{1}{\delta}\big(\epsilon_{j-1} - \epsilon_j\big) - \epsilon_{1,j}(\lambda + \alpha_j(1-\gamma)) + \pi_j\alpha_j\epsilon_2 m \\
\frac{d\epsilon_K}{dt} =& \frac{1}{\delta}\epsilon_{J-1} - \lambda\epsilon_K \\
\frac{d\epsilon_{-2}}{dt} =& \sum_{j=0}^{K}\epsilon_j\alpha_j(1-\gamma) - m\sum_{j=0}^{K} m\pi_j\alpha_j\epsilon_{-2} - \frac{1}{1+\lambda}\epsilon_{-2}
\end{aligned} \tag{3.4}$$

The rest of the proof relies on showing that the dynamical system (**??**) is in fact locally exponentially stable and globally asymptotically stable. We shall do this using lemma

(2) and lemma (3) in the appendix. These proofs are technical and do not add to the exposition provided in the text so we restrict them to the appendix.

As mentioned in the previous section, to complete the proof we need only invoke theorem 1 from Ying (2016). The system therefore converges to the equilibrium of the dynamical system in finite time. □

3.3 Characterizing the policy

If γ denotes the fraction of busy channels, which (under the mean field model) remains to be a constant and is known to a device, then the Bellman equation (3.1) for the discounted problem becomes

$$\begin{aligned} u_i = \max_{\alpha \in [0,A]} \frac{\alpha+\lambda+1/\delta}{A+\lambda+1/\delta}\Bigg(& \frac{(1-\gamma)\alpha}{1/\delta+\lambda+\alpha}\big(R_i+\beta u_{-2}\big) \\ & -\hat{c}(\alpha)\frac{\alpha}{\alpha+\lambda+1/\delta}+\beta\frac{\gamma\alpha}{1/\delta+\lambda+\alpha}u_i+ \\ & \beta\big(\frac{1/\delta}{1/\delta+\lambda+\alpha}u_{i+1}\big)+\beta\big(\frac{\lambda}{1/\delta+\lambda+\alpha}u_0\big)\Bigg). \end{aligned} \tag{3.5}$$

We will henceforth refer to $\hat{c}(\alpha)\frac{\alpha}{\alpha+\lambda+1/\delta}$ as $c(\alpha)$ which obeys all the properties of $\hat{c}(\alpha)$. Note that conditioned on a state transition occurs when the device is in state i, we have the following possibilities:

- With probability $\frac{(1-\gamma)\alpha}{1/\delta+\lambda+\alpha}$, the probing clock ticks and the device finds an idle channel. In this case, the device pays a cost $c(\alpha)$ and receives a reward R_i. The device transits to state 2.

- With probability $\frac{\gamma\alpha}{1/\delta+\lambda+\alpha}$, the probing clock ticks and the device fails to find an idle channel. In this case, the device pays a cost $c(\alpha)$. The device remains in state i.

- With probability $\frac{\delta}{1/\delta+\lambda+\alpha}$, the age of the message increases by one. In this case, the device moves to state $i+1$.

- With probability $\frac{\lambda}{1/\delta+\lambda+\alpha}$, a new message arrives and replace the current message in waiting. In this case, the device moves to state 0.

Note that the term u_i appears at both sides of the Bellman equation, by combing the two terms, we can have (an explicit justification to the equation can be found in the Appendix A):

$$\begin{aligned} u_i = \max_{\alpha \in [0,A]} \frac{1}{1/\delta + \lambda + A - \gamma\beta\alpha} &\Big\{(1-\gamma)\alpha\big(R_i + \beta u_{-2}\big) \\ &-c(\alpha)\left(\alpha + \lambda + 1/\delta\right) + \frac{\beta}{\delta} u_{i+1} + \beta\lambda u_0\Big\} \end{aligned}$$

with the special cases

$$u_{-1} = \beta \frac{\lambda}{A + \lambda + 1/\delta} u_0$$

and

$$u_{-2} = \frac{1+\lambda}{A + \lambda + 1/\delta}\left(r_{-2} + \beta\frac{1}{1+\lambda}u_{-1} + \beta\frac{\lambda}{1+\lambda}u_{-2}\right).$$

Subtracting both sides of the previous equation by $\beta\frac{\lambda}{1+\lambda+A}u_{-2}$, multiplying throughout by $\frac{1+\lambda}{1+\lambda(1-\beta)}$ and substituting u_{-1} in terms of u_0, we obtain

$$u_{-2} = r_{-2}\frac{1+\lambda}{1/\delta + \lambda(1-\beta) + A} + \beta^2 u_0 \frac{\lambda}{1/\delta + \lambda(1-\beta) + A}.$$

Now define

$$r_0 := r_{-2}\frac{1+\lambda}{1/\delta + \lambda(1-\beta) + A}$$

and

$$\eta := \beta\frac{\lambda}{1/\delta + \lambda(1-\beta) + A}$$

which gives us the following expression for u_{-2},

$$u_{-2} = r_0 + \eta\beta u_0$$

Note that we have essentially treated the fraction of busy channels as a constant in studying the Bellman equation above. In other words, a device optimizes its probing strategy assuming γ is fixed. We will justify this later by presenting the conditions under which the system converges to a point mass given by the mean field limit.

Before we proceed, we make the following remarks which will be helpful in later sections,

Remark 2.

1. u_i is bounded below by 0. The lower bound is achieved when we choose $\alpha = \{0, 0, ...\}$, i.e. to do nothing at all no matter the delay, reward or cost.

2. u_i is bounded above by $\frac{R}{1-\beta}$ with $R = R_0 + r_{-2}$.

Now assume we begin by initializing all the devices with the same policy in $\mathcal{P}^K$. From the previous section, it is clear that the fraction of busy channels will converge weakly to some fixed γ. In the rest of the section we will show that the sequence of value functions $\{u_i\}$ is decreasing and so is well defined for all $i \in \{-2, -1, 0, 1, ...\}$. Since both $\{u_i\}$ and $\{\alpha_i\}$ are infinite sequences, we need to establish that $\{\alpha_i\}$ is well defined in the limit as i goes to infinity. (It is not immediate that the map from $\{u_i\}$ to $\{\alpha_i\}$ is sequentially continuous). We use the convergence of $\{u_i\}$ to show that the sequence of $\{\alpha_i\}$ converge to some α_∞. This is followed by bounding the difference in value functions between an optimal policy (which need not lie in $\mathcal{P}^K$) and a policy that lies in $\mathcal{P}^K$ for sufficiently large K. This justifies the mean field model used in the previous section and our proof of convergence.

Proposition 1. *If $\{R_i\}_i$ is a decreasing sequence in i, then sequence $\{u_i\}_i$ is a decreasing sequence in i. Consequently, the sequence converges to some u_∞ in the limit as $i \to \infty$.*

Proof. Let the optimal policy for a device in state i be α_i^* for every i in $\{0, 1, ...\}$ and denote by u_i^* the value functions of the optimal policy. We define function $u_i(\alpha_i)$

as

$$
\begin{aligned}
u_i(\alpha_i) =& \frac{\alpha_i(1-\gamma)}{1/\delta+\lambda+A-\alpha_i\gamma\beta}\big(R_i+\beta u^*_{-2}\big) \\
&- c(\alpha_i)\frac{1/\delta+\lambda+\alpha_i}{1/\delta+\lambda+A-\alpha_i\gamma\beta} \\
&+ \beta\big(\frac{1/\delta}{1/\delta+\lambda+A-\alpha_i\gamma\beta}u^*_{i+1}\big) \\
&+ \beta\big(\frac{\lambda}{1/\delta+\lambda+A-\alpha_i\gamma\beta}u^*_0\big)
\end{aligned}.
$$

From this definition, we have

$$
u_i(\alpha_i^*) = u_i^* = \max_{\alpha} u_i(\alpha),
$$

which implies that

$$
\begin{aligned}
u_i^* &\geq u_i(\alpha^*_{i+1}) \\
&= \frac{\alpha^*_{i+1}(1-\gamma)}{1/\delta+\lambda+\alpha^*_{i+1}(1-\gamma\beta)}\big(R_i+\beta u^*_{-2}\big) \\
&- c(\alpha^*_{i+1})\frac{1/\delta+\lambda+\alpha^*_{i+1}}{1/\delta+\lambda+\alpha^*_{i+1}(1-\gamma\beta)} \\
&+ \beta\big(\frac{1/\delta}{1/\delta+\lambda+\alpha^*_{i+1}(1-\gamma\beta)}u^*_{i+1}\big) \\
&+ \beta\big(\frac{\lambda}{1/\delta+\lambda+\alpha^*_{i+1}(1-\gamma\beta)}u^*_0\big)
\end{aligned}.
$$

Note that we add the superscript $*$ to value function u_i to differentiate the notation from function $u_i(\alpha)$. u_i^* in this proof is the same as u_i defined in (3.5) and the statement of the proposition. Note there is a similarity between $u_i(\alpha^*_{i+1})$ and u^*_{i+1}. Namely,

$$
\begin{aligned}
u_i^* &\geq u_i(\alpha^*_{i+1}) \\
&= u^*_{i+1} + \beta\big(\frac{1/\delta}{1/\delta+\lambda+\alpha^*_{i+1}(1-\gamma\beta)}\big)(u^*_{i+1}-u^*_{i+2})
\end{aligned}.
$$

Rearranging the terms, we get

$$\begin{aligned}
&u_{i+1}^* - u_i^* \\
\leq &\beta\left(\frac{1/\delta}{1/\delta + \lambda + \alpha_{i+1}^*(1-\gamma\beta)}\right)(u_{i+2}^* - u_{i+1}^*) \\
\leq &\beta\big(\frac{1/\delta}{1/\delta + \lambda}\big)(u_{i+2}^* - u_{i+1}^*)
\end{aligned}$$

for all i in $\{0, 1, ...\}$. This yields

$$\left(\beta\big(\frac{1/\delta}{1/\delta + \lambda}\big)\right)^J (u_{i+J+1}^* - u_{i+J+2}^*) \geq u_{i+1}^* - u_i^*.$$

Since the LHS of the inequality above converges to zero as $J \to \infty$, we have $u_i^* \geq u_{i+1}^*$. So, u_i^* is a decreasing sequence which is bounded above and below. Therefore, u_i^* converges to a fixed value. Let $u_\infty^* := \lim_{i\to\infty} u_i$. Then by definition, u_i^* converges to u_∞^*. □

Next we state that α_i^* is a Cauchy sequence that converges to some α_∞.

Lemma 4. *If, R_i is a decreasing sequence in i, then the probing rate under the optimal policy, denoted by α_i^*, is also a decreasing sequence which converges to some α_∞^*.*

The proof of this lemma is presented in Appendix C. We now make a few remarks about the results above.

Remark 3.

1. Both results generalize to an M-player game where the reward and costs are different for different players.

2. It is worth noting that Lemma (4) tells us that a device should probe most aggressively early on and look to drop the packet as the delay increases. Effectively, this observation seems to suggest at steady state the devices will behave such that the packets with the least delay will be more likely to be transmitted before packets with higher delay value.

3. Proposition 1 holds even if the transition from (i) to $(i+1)$ are not exponential but instead are arbitrary, so long as the same process holds for any i and the transition rate only depends on the policy at state i i.e, the rate only depends on α_i.

The next result bounds the difference between a device that uses the optimal policy $\alpha^* = \{\alpha_0^*, \alpha_1^*, ...\alpha_K^*, \alpha_{K+1}^*, ...\}$ and the truncated version $\alpha^{(K)} := \{\alpha_0^*, \alpha_1^*, ...\alpha_K^*, 0, 0, ...\}$ assuming that all other devices choose the same fixed policy $\alpha^{(K)}$. We show that when γ is fixed, given any $\epsilon > 0$, a device can chooses $\alpha^{(K)}$ for a sufficiently large K so that the difference in average utility the device receives comparing with the original version is at most ϵ, where K is independent of the number of devices. Therefore, a device may effectively choose a finite dimensional policy if it wishes to myopically optimize its utility function. This justifies our use of a finite dimensional policy while considering the mean field model to approximate the system.

Proposition 2. *Let α^* and $\alpha^{(K)}$ be as defined above. Given any $\epsilon > 0$, there exists constant β_0 such that for all $\beta < \beta_0$ there exists K large enough so that*

$$|E_X\{u(X, \alpha^*)\} - E_{\tilde{X}}\{u(\tilde{X}, \alpha^K)\}| < \epsilon,$$

where the expectation is taken over the stationary distribution of the device for fixed γ.

The proof of this proposition can be found in Appendix C. Note that we have shown that point-wise convergence of the policy α yields weak convergence of the stationary distribution of a device.

We have now characterized to some extent the mapping from γ to α. Given a fixed γ and a parameter K that is common among all other devices, a device may now use value iteration while fixing $\alpha_k = 0$ for all $k > K$ to arrive at an approximate policy, which is the mapping T_2.

3.4 Existence of MFNE

Based on the results in the previous sections, we will now show that there exists an MFNE using Brouwer's fixed point theorem. We show that both mappings T_1 and T_2 are Lipschitz. Therefore, there exists at least one fixed point (MFNE). This is stated in the theorem below.

Theorem 7. *There exists a constant β_0 such that for any $\beta < \beta_0$, a fixed point for the composite map $T_2 \circ T_1$, denoted by T, exists.*

Proof. We will show in Lemma (11) and Lemma (13) that $T_1 : \alpha \to \gamma$ is Lipschitz and $T_2 : \gamma \to \alpha$ are continuous under the L1-norm. Therefore, the map $T : \alpha \to \alpha$ is also continuous in γ. The policy space, $\mathcal{P}^K$ is clearly convex. Since the domain of γ is compact and T_2 is continuous, so the range of T_2 must be compact. Therefore, by Brouwer's fixed point theorem, there exists a fixed value α^{**} such that:

$$T(\alpha^{**}) = \alpha^{**}.$$

□

The proof of Lemma (11) and Lemma (13) can be found in the appendix.

Having demonstrated that there exists a fixed point, our next theorem shows that the fixed point that is obtained using the map $T_2 \circ T_1$ is in fact an ϵ Nash Equilibrium where ϵ converges to 0 as M and K tend to infinity.

Theorem 8. *The fixed point given by Theorem 7 is an ϵ-Nash equilibrium when the set of available policies are from $\mathcal{P}^K$, with $\epsilon \to 0$ as M tends to infinity.*

Proof. The main idea of the proof follows from the fact if one player chooses to deviate by choosing any policy in $\mathcal{P}^K$, then the mean field, γ changes by at most ϵ where ϵ goes to zero as M tends to infinity. This follows from a simple extension of the proof of ϵ Nash Equilibrium in Narasimha *et al.* (2020). The proof relies on Stein's method for finite state Markov chains.

Now, since the mean field γ only deviates by a small amount and $T_2 : \gamma \to \alpha$ is continuous, the best response policy in $\mathcal{P}^K$ will lie close to the mean field policy.

□

Here we would like to make few comments:

Remark 4.

1. Theorem 8 does not rely on asymptotic independence of the devices in our system. Therefore, it is not trivial to show that if a finite set of players deviate the mean field remains unchanged. Infact under typical law of large numbers results the order of ϵ is $\mathrm{O}(\sqrt{M})$ but in our case ϵ will actually be $\mathrm{O}(M^{\frac{1}{3}})$.

2. Since $\mathcal{P}^K$ is ϵ close to an optimal policy for fixed γ one might be tempted to state that the result of Theorem 8 holds for any policy instead of policies in $\mathcal{P}^K$. This is not straightforward since if a single node deviates with an arbitrary policy, the corresponding fraction of busy channels need not deviate by ϵ. Therefore, the best response need not necessarily be the MFNE policy.

3. While our theorems do not limit the number of fixed points, we strongly believe that there is a unique fixed point, primarily because we believe that the function T is decreasing in γ. Since our fixed points are the set of all γ such that $\gamma = T(\gamma)$ this would ensure that the fixed point is unique. This conjecture of a unique MFNE is a topic to be investigated later.

From Finite to Countable State MFNE

For each K we know that there exists a MFNE solution to our problem. Let α^{*K} denote such a solution. This solution corresponds to a mean field distribution

$\pi^K = \{\pi_{-2}, \pi_{-1}, \pi_0, ...\pi_{K-1}, S_K\}$ on the state space, $\{-2, -1, 0, 1, ..., K-1, K\}$. Our results state that the empirical distribution will converge to an invariant measure on this state space. We can treat both the state space and the policy space α as vectors belonging to a finite dimensional subspace inside an infinite dimensional space i.e, $\alpha^K = \{\alpha_0, \alpha_1, ..\alpha_{K-1}, 0, 0...0\}$ and $\{-2, -1, 0, 1, ..., K-1, K, 0, 0...\}$. We will now imbue the infinite dimensional state and policy space with the *topology of point-wise convergence.*

Note, this is a very natural formalization of our work so far and is closely related to our finite dimensional approximation of our policy, Proposition (2). With this technical manipulation, both the state space and the policy space are now separable metric spaces. Let K_1 be the smallest integer such that α^{*K_1} is ϵ close to the MFNE policy. $\{K_1, K_2, ...\}$ is an increasing sequence of integers. From our previous results we know that for each $K_i \in \mathbb{N}$ in this sequence, at least one fixed point exists in $\mathcal{P}^{K_i}$ with the corresponding mean field limit π^{K_i}. We now have the tools to state our theorem that allows us to look at the countably infinite system.

Theorem 9. *The sequence, $\{\pi^{K_i}\}$ has a weakly convergent sub-sequence $\{\pi^{K_{j(i)}}\}$ such that $j(i) \to \infty$ as $i \to \infty$ and $\{\pi^{K_{j(i)}}\} \to \pi^*$ where π^* is given by the fixed point of the system of ODEs:*

$$\begin{aligned}
\frac{dq_{-1}}{dt} =& -\lambda q_{-1} + \frac{1}{1+\lambda} q_{-2} \\
\frac{dq_{-2}}{dt} =& \sum_{i=0}^{\infty} (1-\gamma)\alpha_i q_i - \frac{1}{1+\lambda} q_{-2} \\
\frac{dq_0}{dt} =& \lambda\big(1 - q_0 - q_{-2}\big) - q_0 \frac{1}{\delta} - \alpha_0(1-\gamma) q_0 \\
\frac{dq_j}{dt} =& \big(q_{j-1} - q_j\big)\frac{1}{\delta} - \alpha_j(1-\gamma)q_j - \lambda q_j
\end{aligned}$$

and $\alpha = \{\alpha_0, \alpha_1 ...\}$, is the solution to the countably infinite state MDP for fixed $\gamma = \pi^*_{-2}$

The theorem above states that a fixed point exists and solves the MDP for the infninite state MDP problem that was described in section (3.1). The proof is a simple application of Prokhorov's theorem and can be found in the Appendix D. Here we will make some important remarks,

Remark 5.

1. While we have proved the existence of a fixed point for the countably infinite system, it is still an open problem to show that this fixed point is indeed a Nash equilibrium. This is primarily because extending the mean field result to a general countably infinite CTMC is still an open problem. Therefore, if a single user chooses to deviate from the MFNE policy in this case, we do not know of any way to bound the deviation of the mean field system.

2. In keeping with Remark (3), part (3), if T is indeed a decreasing function in γ then; it might be possible to extend the results further and show that the sequence, $\{\pi^{K_i}\}$ is in fact a Cauchy sequence which is convergent instead of dealing with sub-sequences.

3.5 Implementation

The previous section proved that there exists at least one fixed point and the fixed point obtained is a local ϵ Nash Equilibrium. However, in the absence of contraction maps it is difficult to imagine how the device may achieve these equilibria. Here we propose a scheme by which a device may achieve this equilibria.

We first note that while α can in general be a complicated variable even when it belongs to $\mathcal{P}^K$ the variable γ is a one dimensional real variable restricted to a closed bounded set, $(0, 1)$. For a fixed γ a device may use policy iteration to find a policy α in $\mathcal{P}^K$ that is ϵ close to the optimal policy and for this policy α explicitly compute

$T_2(\alpha)$. This gives the device an estimated value of $T(\gamma)$ for a fixed value of γ. The device can now repeat this process for n such values of γ and use this to estimate the function T computationally. The parameter n can be chosen by the designer and is independent of K or the number of devices. Now the device can find fixed points by solving $\gamma = T(\gamma)$. The device now picks the fixed point with the highest expected utility with ties broken based on lower fraction of busy channels. (Although we strongly believe there will only be one fixed point). [H] Algorithm to find fixed points **input** : Rewards $(R_j)_j$, $c(.)$, β, r_0, η, λ, δ

initialization : pick γ_i uniformly from the interval $(0, 1)$

$i < n$ $\alpha \leftarrow$ policy iteration(γ, $(R_j)_j$, $c(.)$, β, r_0, η, δ, λ) $T(i) \leftarrow \frac{1-\theta(\alpha)}{\kappa-\theta(\alpha)}$ i $\leftarrow$ i + 1

interpolate T ; find γ such that $\gamma = T(\gamma)$ **output** : γ Ostensibly one can view this algorithm as the device playing the game with itself in its own head to estimate the LMFNE. Under these conditions the LMFNE assumptions of infinite players are justified and consistent thus, leading to local Nash equilibria for the devices. Since the devices are homogeneous they will pick the same policies achieving the computed γ. An example implementation can be found in Figure 3.2.

3.6 Simulation Results

We present simulation results to demonstrate the performance of the proposed algorithm. We consider the following setting: $M = N$, i.e. the number of channels is equal to the number of devices, $K = 25$, $R_i = 10 \times 2^{-i}$, $\beta = 0.1$, $\delta = 1$, $A = 5$, and $c(\alpha) = 10\alpha^2$. We evaluate the delay experienced per packet when the system reaches MFNE and compare it to the throughput-oriented MAC protocol proposed in Narasimha *et al.* (2020). The protocols are:

- AD-MAC : This is our age-dependent distributed MAC. We varied the arrival rate λ from 0 to 2.

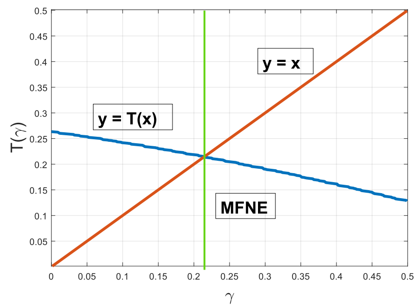


Figure 3.2: An example implementation with $K = 25$, $R_i = 2^{-i}$, $\beta = 0.1$, $\delta = 1$, $\lambda = 0.5$ and $c(\alpha) = 0.3\alpha^2$

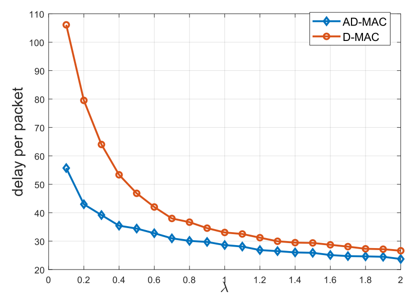


Figure 3.3: A comparison of the delays experienced per packet delivered over some time duration.

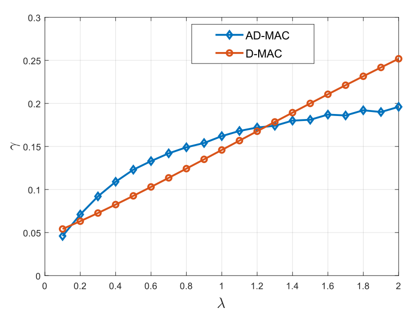


Figure 3.4: Comparing γ as a function of λ for the two protocols.

- D-MAC : This is the distributed MAC protocol proposed in Narasimha *et al.* (2020). We chose c to be 10 and evaluated the delay over the same range of λ. Our choices of c, λ in this case ensure that the MFNE exists as required in Narasimha *et al.* (2020).

The per packet delays are shown in Figure 3.3, where we can observe that AD-MAC has significant smaller per packet delay when the arrival rate is low, and the delay of AD-MAC is always smaller than that of D-MAC in our simulations. In addition to the delay, we also compared the fraction of occupied channels, which reflects the system throughput. As we can observe from Figure 3.4, AD-MAC achieves higher throughput when $\lambda \leq 1.2$ (with smaller per packet delay as well). For $\lambda > 1.2$, the throughput is lower than that under D-MAC. This loss is to achieve lower per-packet delay.

Chapter 4

STRATEGIC DECONFLICTION

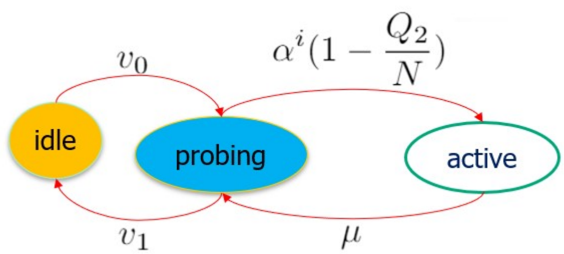
4.1 System Model

We consider a disaster monitoring system where there are M regions to be monitored by N UAVs. For now we assume that the regions themselves are equal in size and the UAVs are homogeneous, we will revisit this assumption in the simulations. In our setting we would like to consider the case when N is large but M/N remains constant. This is a common theme in an urban setting where traffic management becomes problematic as the number of UAVs grow large.

Our main problem is to formulate a protocol that allows UAVs to be assigned tasks towards monitoring a disaster region in a distributed fashion. We assume that the UAVs have a charging/ repair station the UTM operator may use for repairing, charging the UAV or for routine maintenance. We will call this state the idle state. As the UAVs are charged they migrate to a region where they may interact with some central base station or monitoring network. This station has a list of un-monitored regions. We also assume that the central station is the primary network that the UAV may use to send data to the UTM operator or other monitoring stations. When the devices are capable of interacting with the base station they probe it for un-monitored regions at a rate that they may control. For UAV i we let the exponential probing rate (by design) be given by $\alpha^{(i)}$. We call this state the probing state. The UAV's drop off from the probing state for recharging or repair to the idle state as an exponential random variable with parameter v_0. We model the time taken to move from idle state to the probing state as another exponential random variable with rate v_1.

When a UAV finds a region to monitor, it travels to this region and begins to monitor it. Given the UAV spends some time t in this region it does not know how much more time to spend in this region. For example, it may be looking for an event in a disaster region whose occurrence time is unknown. We therefore model the monitoring time by some exponential time with rate μ.

We now have a continuous time markov chain model for each UAV. We show the corresponding state space evolution below. Note the continuous time markov chain is reversible and hence, a stationary distribution for this system always exists. We provide a diagram for the evolution of the state space below.



Note, the transition rate of the markov process from state probing to monitoring depends on the fraction of UAVs currently monitoring the region. This makes our markov process a density dependent markov process.

We will summarize the UAV model below :

- There are three states for each UAV, idle, probing and monitoring (denoted, $0, 1, 2$ respectively).
- Let $Q_0(t)$ be the total number of UAVs currently in the idle state, Q_1 the corresponding number of UAVs in the probing state and Q_2 the number of UAVs monitoring an area.
- A UAV goes from the idle state to the probing state at rate v_0 and returns from the probing state to the idle state at rate v_1.

- Similarly, a UAV probes a region to check if it is being monitored uniformly at random at rate $\alpha^{(i)}$. Therefore, its rate of finding an un-monitored region is $\alpha^{(i)}(1-\frac{Q_2}{N})$. A UAV monitors a region for an exponential amount of time with average μ before returning to the probing state.

Compounded with the problem of density dependence each UAV is treated as an individual agent that wishes to optimize its own long term time average utility function. We model the long term utility of each device in terms of the fraction of time the UAV will spend in a monitoring region minus the average cost of probing the central base station which we model as a convex cost function depending on the probing rate and time spent by the device probing for regions.

If π_j^i is the fraction of time spent by the UAV $i \in \{1, 2...N\}$ in state $j \in \{0, 1, 2\}$ then the utility function for each UAV is:

$$J(\alpha^{(i)}) := \pi_2^i - c(\alpha^{(i)}\pi_1^i) \tag{4.1}$$

Where c is any twice differentiable, increasing, convex cost function.

Note, while this function does not explicitly depend on the density of devices in any given state, the values of π_j^i will implicitly depend on these density values. In general this density could be a time varying function which makes this problem extremely difficult since that would imply that optimizing J over time will yield a time varying optimal policy. If formulated as a Bayesian game the problem quickly becomes intractable to solve the optimal policy for any given device. With this in mind we turn towards formulating this problem as a mean field game!

4.2 Mean Field Limit

In this section we consider the mean field limit as described previously, using the detail balanced equation:

$$\dot{q_0} = -q_0 v_0 + q_1 v_1$$
$$\dot{q_1} = q_0 v_0 - q_1 v_1 + \mu q_2 - q_1 \alpha(1-\gamma)$$
$$\dot{q_2} = -\mu q_2 + q_1 \alpha(1-\gamma)$$

Let $\{\pi_0, \pi_1, \pi_2\}$ be the fixed points of the system above. Note, they are given by the following set of equations,

$$\pi_0 = \pi_1 \frac{v_1}{v_0}$$
$$\pi_2 = \pi_1 \alpha \frac{(1-\gamma)}{\mu}$$
$$\pi_0 + \pi_1 + \pi_2 = 1$$

We must first begin by showing that the equations above yield a unique solution to the problem. We will do so below,

Using the consistency equation $q_2(t) = \gamma(t)$ and the equation above in π_2, we get:

$$\pi_2 = \frac{\pi_1 \alpha}{\mu + \alpha \pi_1}$$

Combining the equations and writing them in terms of π_1 we get,

$$\pi_1 \left(1 + \frac{v_1}{v_0} + \frac{\alpha}{\mu + \alpha \pi_1} \right) = 1 \tag{4.2}$$

The equation above is of the form $f(q_1) = 1$. It can be checked that the solution to this equation is unique because f is an increasing function in q_1. Let $Q_i(\infty)$ be the number of UAVs in state i as time tends to infinity, i.e, the steady state of the system, the next theorem formally says the Markovian system $\{\frac{Q_0(\infty)}{N}, \frac{Q_1(\infty)}{N}, \frac{Q_2(\infty)}{N}\}$ converges weakly to the fixed point desribed above,

Theorem 10. *Suppose every device fixes a polic α, the system, $\{\frac{Q_0(\infty)}{N}, \frac{Q_1(\infty)}{N}, \frac{Q_2(\infty)}{N}\}$ converges weakly to the fixed point $\{\pi_0, \pi_1, \pi_2\}$ as N tends to infinity. This now defines our map T_1.*

Proof. The proof is an application of Theorem 1 in [22]. The theorem states five conditions that are sufficient to guarantee the weak convergence to the fixed point of the mean field model. We next verify these conditions under our model:

- Bounded transition rate: This condition can easily be verified from the system model. At any point in time, the rate of transition from any state to any other state is bounded above by $v_1 + v_0 + \mu + \alpha$.

- Bounded state transition condition: Since our model is a collection of M CTMCs whose transition rates are determined by exponential clocks, at most one transition can occur at a time. Therefore, the state transitions are bounded.

- Perfect Mean Field Model: Using the system model it can be checked that the equations (2) are derived from the detailed balance equations.

- Partial Derivative condition: It can be checked that the partial derivatives for the system (2) exist and are Lipschitz.

- Stability conditions: The global asymptotic convergence and local exponential convergence can be found using the Lyapunov function

$$V(t) = |q_0(t) - \pi_0| + |q_1(t) - \pi_1| + |q_2(t) - \pi_2|$$

  and the analysis from the 2nd Chapter.

□

4.3 Existence and Uniqueness of Mean Field Equilibrium

Let γ be the fraction of regions being monitored. We will assume here that the UAVs know this value of γ. Suppose each UAV chooses to optimize its policy as a best response to the value of γ, due to homogeneity, each UAV will choose the same policy as its best response function. For fixed γ the UAV must optimize,

$$J(\alpha, \gamma) = \pi_2(\alpha, \gamma) - c(\pi_1(\alpha, \gamma)\alpha)$$

Lemma 5. *When γ is fixed, the devices will pick an optimal policy $\tilde{\alpha}$ given by the unique solution of the following equation,*

$$\alpha\pi_1(\alpha, \gamma) = c'^{-1}\left(\frac{1-\gamma}{\mu}\right) \tag{4.3}$$

Proof. Note, for any convex, twice differentiable, increasing function c, if we fix γ, the best response function given by,

$$\tilde{\alpha} =_{\alpha} J(\alpha, \gamma)$$

is continuous in γ and can be computed simply by taking the derivative with respect to α and setting it to 0. Note, each agent treats γ like a constant that it cannot influence. Therefore, when taking the derivative, we will do the same, one can see this as the fast time scale action from assumption 2. From the previous section,

$$\pi_1 = \frac{1}{1 + \frac{v_1}{v_0} + \alpha\frac{1-\gamma}{\mu}}$$

and

$$\pi_2 = \pi_1 \frac{\alpha(1-\gamma)}{\mu}$$

Note, the derivative of π_2 w.r.t α is:

$$\frac{d\pi_2}{d\alpha} = \frac{1-\gamma}{\mu}\left(\pi_1 + \alpha\frac{\pi_1}{d\alpha}\right)$$

The derivative of $c(\alpha\pi_1)$ is

$$\frac{dc(\alpha\pi_1)}{d\alpha} = c'(\alpha\pi_1)\left(\pi_1 + \alpha\frac{\pi_1}{d\alpha}\right)$$

which can be rewritten in terms of $\frac{d\pi_2}{d\alpha}$, note this term never goes to zero.

$$\frac{dc(\alpha\pi_1)}{d\alpha} = c'(\alpha\pi_1)\frac{d\pi_2}{d\alpha}\frac{\mu}{1-\gamma}$$

Also note, since c is an increasing function and convex, $c'(\alpha\pi_1) > 0$. Subsitituting this in the derivative of J with respect to α we get, at $\alpha = \tilde{\alpha}$

$$\frac{d\pi_2}{d\alpha}\left(1 - \frac{\mu}{1-\gamma}c'(\alpha\pi_1)\right) = 0$$

which means at $\alpha = \tilde{\alpha}$,

$$1 = \frac{\mu}{1-\gamma}c'(\alpha\pi_1)$$

For fixed γ, $\alpha\pi_1$ is an increasing function in α and since c is convex, c' is an increasing function. Therefore, the point where $c'(\alpha\pi_1)$ meets $\frac{1-\gamma}{\mu}$ is unique. This completes the proof. $\square$

Thus, for any fixed γ, the devices will choose the same policy given by (4.3). We will denote (4.3) by the map $T_2 : \gamma \to \tilde{\alpha}$.

Given an initial policy that every UAV follows, the system will converge weakly to a fixed γ by theorem (10), given by the map T_1. Given this γ, the devices (due to homogeneity) will pick the same policy to improve their utility functions $\tilde{\alpha}$, given by (4.3), the map T_2. Thus, the composition of the maps, $T_2 \circ T_1$ takes a point in the policy space α and maps it to another point in the policy space $\tilde{\alpha}$. We may now ask if a mean field equilibrium exists in this case. The next theorem answers this question to the positive.

Theorem 11. *Under assumption 1 and 2, there exists a unique mean field equilibrium, α^* such that,*

$$\alpha^* = T_2 \circ T_1(\alpha^*) \tag{4.4}$$

Proof. We will use Schauder's fixed point theorem to prove this result, for completeness we state the theorem here,

Theorem 12. *Let A be a closed convex subset of a Banach space and assume there exists a continuous map Tsending A to a countably compact subset $T(A)$ of A. Then T has fixed points.*

Note, it is easy to check that both map $T_1 : \alpha \to \gamma$ and $T_2 : \gamma \to \alpha$ are continuous. Since, γ lies between 0 to 1, it lies in a relatively compact set. Therefore, the closure of $T_1(0, 1)$ is compact. Since, γ takes values from a continuous interval in the real line $\mathbb{R}$ and the set of policies are also from the real line, the set of values of α must also belong to an interval in the real line. Therefore, the map $T_2 \circ T_1(\alpha)$ maps the policies to a convex, relatively compact set. Compactness in this case automatically implies countably compact, hence, by Schauder's fixed point theorem there exists at least one fixed point α^* such that $T_2 \circ T_1(\alpha^*) = \alpha^*$.
It is easy to check that $T_1 : \alpha \to \gamma$ is increasing in α, that is, if α increases then so does γ. The next lemma describes the map T_2 and is left as an exercise to the reader.

Lemma 6. *Map $T_2 : \gamma \to \alpha$ (which as a reminder occurs in the slow time scale) is decreasing in γ*

Therefore, the composition of T_1 and T_2 is decreasing. Note, the mean field equilibrium is a solution of $\alpha^* = T_2 \circ T_1(\alpha^*)$, therefore, it is the intersection of the curve, $y = T_2 \circ T_1(x)$ with the curve, $y = x$. The first curve was shown to be decreasing, the second curve is increasing, therefore, they can intersect at only one place. $\square$

Finally, as in the previous chapters can we relate the mean field equilibrium that we found to the notion of a Nash equilibrium. The next theorem states that the unilateral deviation of a single agent doesn't change the map T_1 by a factor greater that $\frac{1}{N^{1/3}}$. This means that as the number of UAVs grow large, the map remains intact, therefore,

Theorem 13. *Every mean field equilibrium found above is an ϵ- Nash equilibrium with $\epsilon = \frac{1}{N^{1/3}}$. That is, no UAV can benefit more than a factor of $\frac{1}{N^{1/3}}$ by choosing to deviate from the mean field equilibrium.*

Remark 6.

- Since both maps are Lipschitz, one can look for the region where the composition of the maps yields a Lipschitz constant less than 1 as in Chapter 2. In this case one can refer to Chapter 2 for the simple implementation of the a protocol where the UAVs will perform policy improvement on the slower timescale to arrive at the Nash equilibrium.

- However, for the more general case, one may use the algorithm from Chapter 3 where the UAV essentially plays the game in its head to find the Nash equilibrium. Due to the uniqueness of equilibrium we find that all the UAVs will arrive at the same policy at the end of the game.

REFERENCES

"Mean field games, web-page by terry tao", https://terrytao.wordpress.com/2010/01/07/mean-field-equations/ (2010).

Carmona, G., "Nash equilibria of games with a continuum of players, obtainable at https://ssrn.com/abstract=882468 or http://dx.doi.org/10.2139/ssrn.882468", (2004).

Dawson, V., D., "J. stochastic mckean-vlasov equations", URL https://doi.org/10.1007/BF01295311 (1995).

Evans, L. C., *Partial differential equations* (American Mathematical Society, Providence, R.I., 2010).

Gast, N., "Expected values estimated via mean-field approximation are 1/n-accurate", Proc. ACM Meas. Anal. Comput. Syst. **1**, 1, 17:1–17:26 (2017).

Gast, N. and B. Van Houdt, "A refined mean field approximation", in "Proc. Ann. ACM SIGMETRICS Conf.", (Irvine, CA, 2018).

Iyer, K., R. Johari and M. Sundararajan, "Mean field equilibria of dynamic auctions with learning", Management Science **60**, 12, 2949–2970 (2014).

Kam, C., S. Kompella, G. D. Nguyen, J. E. Wieselthier and A. Ephremides, "On the age of information with packet deadlines", IEEE Transactions on Information Theory **64**, 9, 6419–6428 (2018).

Kaul, S., M. Gruteser, V. Rai and J. Kenney, "Minimizing age of information in vehicular networks", in "IEEE Int. Conf. Sensor, Mesh and Ad Hoc Communications and Networks (SECON)", pp. 350–358 (2011).

Kaul, S., R. Yates and M. Gruteser, "Real-time status: How often should one update?", in "Proc. IEEE Int. Conf. Computer Communications (INFOCOM)", pp. 2731–2735 (2012).

Khalil, H. K., *Nonlinear systems* (Prentice Hall, 2001).

Lasry, J.-M. and P.-L. Lions, "Mean field games", Japanese Journal of Mathematics **2**, 1, 229–260 (2007).

Li, J., R. Bhattacharyya, S. Paul, S. Shakkottai and V. Subramanian, "Incentivizing sharing in realtime D2D streaming networks: A mean field game perspective", IEEE/ACM Trans. Netw. (2016).

Narasimha, D., S. Shakkottai and L. Ying, "A mean field game analysis of distributed MAC in ultra-dense multichannel wireless networks", IEEE/ACM Transactions on Networking pp. 1–14 (2020).

Sun, Y., I. Kadota, R. Talak and E. Modiano, "Age of information: A new metric for information freshness", Synthesis Lectures on Communication Networks **12**, 2, 1–224 (2019).

Sznitman, A.-S., "Topics in propagation of chaos", in "Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989", edited by P.-L. Hennequin, pp. 165–251 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1991).

Xia, B., S. Shakkottai and V. Subramanian, "Small-scale markets for bilateral resource trading in the sharing economy", Proc. IEEE Int. Conf. Computer Communications (INFOCOM) (2018).

Yates, R. D. and S. K. Kaul, "The age of information: Real-time status updating by multiple sources", IEEE Transactions on Information Theory **65**, 3, 1807–1827 (2018).

Ying, L., "On the approximation error of mean-field models", in "Proc. Ann. ACM SIGMETRICS Conf.", (Antibes Juan-les-Pins, France, 2016).

Ying, L., "Stein's method for mean field approximations in light and heavy traffic regimes", Proc. ACM Meas. Anal. Comput. Syst. **1**, 1, 12:1–12:27 (2017).

Yong, J. and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations* (Springer, 2010).

APPENDIX A

PROOFS FOR DISTRIBUTED MAC

A.1 Proof of Theorem 1

To understand the mean-field model (2.4), consider $Q_1(t)$ and a sufficiently small time interval δ. According to a standard argument of continuous-time Markov chain, we have

$$\begin{aligned} & E[Q_1(t+\delta) - Q_1(t)|\mathbf{Q}(t) = \mathbf{Q}] \\ = \; & \lambda Q_0\delta - k\left(1 - \left(\frac{Q_2}{N}\right)^{d/k}\right) Q_1\delta + O(\delta^2), \end{aligned} \tag{A.1}$$

where $\lambda Q_0\delta$ is the probability that during $[t, t+\delta]$, one of the devices moves from idle to probing, and $k\left(1 - \left(\frac{Q_2}{N}\right)^{d/k}\right) Q_1\delta$ is the probability that during $[t, t+\delta]$, one of the devices moves from probing to transmitting. Now dividing $M\delta$ on both sides, we obtain

$$\begin{aligned} & \frac{E[Q_1(t+\delta) - Q_1(t)|\mathbf{Q}(t) = \mathbf{Q}]}{M\delta} \\ = \; & \lambda\frac{Q_0}{M} - k\left(1 - \left(\frac{mQ_2}{M}\right)^{d/k}\right) \frac{Q_1}{M} + O(\frac{\delta}{M}), \end{aligned} \tag{A.2}$$

Now defining $q_i = \frac{Q_i}{M}$ and

$$\dot{q}_1 = \lim_{\delta\to 0} \frac{E[Q_1(t+\delta) - Q_1(t)|\mathbf{Q}(t) = \mathbf{Q}]}{M\delta},$$

we have

$$\dot{q}_1 = \lambda q_0 - k(1 - (mq_2)^{d/k})q_1, \tag{A.3}$$

i.e. the mean-field model for q_1. The rest of the mean-field model can be similarly obtained. We can see that the mean-field model approximates the original stochastic system by using the expected drift (A.2) as the system dynamic.

In the following lemma, we first show that mean-field model (2.4) has a unique equilibrium.

Lemma 7. *Given $d > 0$ and $k > 0$, mean field model (2.4) has a unique equilibrium.*

Proof. The equilibrium point of mean field model (2.1) satisfies the following fix point equations:

$$\begin{aligned} q_0^* &= \frac{k(1 - (mq_2^*)^{d/k})}{\lambda(1 + k(1 - (mq_2^*)^{d/k})(1 + \lambda + \frac{1}{\lambda}))} \\ q_1^* &= \frac{1}{(1 + k(1 - (mq_2^*)^{d/k})(1 + \lambda + \frac{1}{\lambda}))} \quad . \\ q_2^* &= \frac{(1+\lambda)k(1 - (mq_2^*)^{d/k})}{(1 + k(1 - (mq_2^*)^{d/k})(1 + \lambda + \frac{1}{\lambda}))} \end{aligned}$$

Note that q_0^* and q_1^* are uniquely determined by q_2^*. Therefore, we next show that q_2^* has a unique solution. Recall that $\gamma^* = mq_2^*$. Substituting it into the third equation above, we have

$$\gamma^* = m\frac{(1+\lambda)k(1-(\gamma^*)^{d/k})}{(1+k(1-(\gamma^*)^{d/k})(1+\lambda+\frac{1}{\lambda}))}. \tag{A.4}$$

Define function $\theta(\cdot)$ such that

$$\theta(\gamma^*) = \frac{m(1+\lambda)k(1-(\gamma^*)^{\frac{d}{k}})}{1+k(1-(\gamma^*)^{\frac{d}{k}})((1+\lambda)+\frac{1}{\lambda})}$$

Notice that $\theta(\gamma^*)$ is monotonically decreasing function in γ^* because

$$\begin{aligned}\frac{d\theta(\gamma^*)}{d\gamma^*} =& -m(1+\lambda)d(\gamma^*)^{d/k-1}\times \\ & \left(\frac{1}{\left(1+k(1-(\gamma^*)^{d/k})((1+\lambda)+\frac{1}{\lambda})\right)^2}\right) \\ <&0.\end{aligned}$$

Further note that

$$\theta(0) = \frac{m(1+\lambda)k}{\left(1+k((1+\lambda)+\frac{1}{\lambda})\right)} > 0$$

and

$$\theta(1) = 0.$$

Since γ^* is strictly increasing in γ^* and $\theta(\gamma^*)$ is strictly decreasing in γ^*, we conclude that $\gamma^* = \theta(\gamma^*)$ has a unique solution, which concludes the lemma. $\square$

We need to verify the conditions laid out in the main result of Ying (2016) in order to complete our proof of convergence. It is straightforward to see that the system has a bounded transition rate and only a constant number of devices can go from one state to another at any given point in time t since they are all running exponential clocks. The ODE systems we use are all twice differentiable. The perfect mean field model condition,can be verified when $k = d$ through the preceding sections and equation (A.2) - (A.3). Recall each device uses policy (d, d). The mean field model under this policy is similar to Equation 2.4 and is the following nonlinear system:

$$\begin{aligned}\frac{dq_0}{dt} &= -\lambda q_0 + \frac{1}{1+\lambda}q_2 \\ \frac{dq_1}{dt} &= \lambda q_0 - d(1-mq_2)q_1 \\ \frac{dq_2}{dt} &= d(1-mq_2)q_1 - \frac{1}{1+\lambda}q_2\end{aligned} \tag{A.5}$$

Let (q_0^*, q_1^*, q_2^*) denote the unique equilibrium point of this dynamical system. The uniqueness of the equilibrium point is due to Lemma (7). Define $\epsilon_i(t)$ to be

$$\epsilon_i(t) = q_i(t) - q_i^*.$$

Then the dynamical system (A.5) can be equivalently represented by :

$$\begin{aligned}\frac{d\epsilon_0}{dt} &= -\lambda\epsilon_0 + \frac{1}{1+\lambda}\epsilon_2 \\ \frac{d\epsilon_1}{dt} &= \lambda\epsilon_0 - d(1-mq_2^*)\epsilon_1 + mdq_1\epsilon_2 \\ \frac{d\epsilon_2}{dt} &= d(1-mq_2^*)\epsilon_1 - mkq_1\epsilon_2 - \frac{1}{1+\lambda}\epsilon_2\end{aligned} \tag{A.6}$$

It is clear from the definition that $\sum_{i\in 0,1,2}\epsilon_i = 0$ for any time t. The final condition to verify before using main result in Ying (2016) is to check the dynamical system is globally asymptotically stable and locally exponentially stable. We proceed to show this in the following lemma.

A.2 Proof of Stability Properties

Lemma 8. *The dynamical system described by (A.6) is asymptotically stable for any valid ϵ_i and locally exponentially stable near the origin.*

Proof. We prove the first part of the lemma using the Lyapunov theorem Khalil (2001). Define Lyapunov function $V(\epsilon)$ such that

$$V(\epsilon) = |\epsilon_0| + |\epsilon_1| + |\epsilon_2|. \tag{A.7}$$

Note that $\sum_i \epsilon_i(t) = 0$ for all t, so at least one of the ϵ_i is negative and one is positive when $\epsilon \neq 0$.

We first analyze the cases where only one ϵ_i is strictly negative, which includes the following three cases.

Case I: $\epsilon_0 < 0$, $\epsilon_2 \geq 0$ and $\epsilon_1 \geq 0$

In this case, we have

$$V(\epsilon) = -\epsilon_0 + \epsilon_1 + \epsilon_2.$$

Therefore,

$$\begin{aligned}\frac{dV}{dt} =& -\frac{d\epsilon_0}{dt} + \frac{d\epsilon_1}{dt} + \frac{d\epsilon_2}{dt} \\ =&\lambda\epsilon_0 - \frac{1}{1+\lambda}\epsilon_2 + \lambda\epsilon_0 - k(1-mq_2^*)\epsilon_1 \\ &+ md\epsilon_2 q_1 + d(1-mq_2^*)\epsilon_1 - md\epsilon_2 q_1 - \frac{1}{1+\lambda}\epsilon_2 \\ =& -2\frac{d\epsilon_0}{dt} \\ =&2\lambda\epsilon_0 - 2\frac{1}{1+\lambda}\epsilon_2 < 0.\end{aligned}$$

Case II: $\epsilon_1 < 0$, $\epsilon_0 > 0$, **and** $\epsilon_2 \geq 0$

In this case, we have

$$V(\epsilon) = \epsilon_0 - \epsilon_1 + \epsilon_2.$$

Therefore,

$$\begin{aligned}\frac{dV}{dt} =& \frac{d\epsilon_0}{dt} - \frac{d\epsilon_1}{dt} + \frac{d\epsilon_2}{dt} = -2\frac{d\epsilon_1}{dt} \\ =& -2\lambda\epsilon_0 + 2d(1 - mq_2^*)\epsilon_1 - 2mdq_1\epsilon_2 < 0.\end{aligned}$$

Case III: $\epsilon_2 < 0$, $\epsilon_0 > 0$, **and** $\epsilon_1 \geq 0$

In this case, we have

$$V(\epsilon) = \epsilon_0 + \epsilon_1 - \epsilon_2.$$

Therefore,

$$\begin{aligned}\frac{dV}{dt} =& \frac{d\epsilon_0}{dt} + \frac{d\epsilon_1}{dt} - \frac{d\epsilon_2}{dt} = -2\frac{d\epsilon_2}{dt} \\ =& -2d(1 - mq_2^*)\epsilon_1 + \frac{1}{1+\lambda}(2mdq_1 + 2)\epsilon_2 < 0.\end{aligned}$$

For the cases where one ϵ_i is strictly positive, we can similarly show $\frac{dV}{dt} < 0$. For example, when $\epsilon_0 > 0$, $\epsilon_2 \leq 0$ and $\epsilon_1 \leq 0$, following a similar analysis to Case I, we have

$$\frac{dV}{dt} = 2\frac{d\epsilon_0}{dt} = -2\lambda\epsilon_0 + 2\frac{1}{1+\lambda}\epsilon_2 < 0.$$

Therefore, based on the Lyapunov theorem, we conclude that the system is asymptotically stable.

To prove that the system is locally exponentially stable, we need to show that the linearized system matrix around its equilibrium is negative definite, i.e, has strictly negative eigenvalues. The linearized dynamical system is given by:

$$\begin{aligned}\frac{d\epsilon_0}{dt} &= -\lambda\epsilon_0 + \frac{1}{1+\lambda}\epsilon_2 \\ \frac{d\epsilon_2}{dt} &= \left(-d(1 - mq_2^*) - mdq_1^* - \frac{1}{1+\lambda}\right)\epsilon_2 - d(1 - mq_2^*)\epsilon_0\end{aligned}, \tag{A.8}$$

where we used the fact $\epsilon_1 = -\epsilon_0 - \epsilon_2$ and eliminated one of the equations from the dynamical system.

The matrix corresponding to the linearized form can be written as:

$$A = \begin{bmatrix} -\lambda & \frac{1}{1+\lambda} \\ -d(1 - mq_2^*) & -d(1 - mq_2^*) - mdq_1^* - \frac{1}{1+\lambda} \end{bmatrix}$$

Let η be an eigenvalue of A, Then η must satisfy

$$(-\lambda - \eta)\big(- d(1 - mq_2^*) - mdq_1^* - \frac{1}{1+\lambda} - \eta\big) \\ + \frac{d}{1+\lambda}(1 - mq_2^*) = 0.$$

If $\eta \geq 0$, then the first term is strictly positive and

$$(-\lambda - \eta)\big(- d(1 - mq_2^*) - mdq_1^* - \frac{1}{1+\lambda} - \eta\big) \\ + \frac{d}{1+\lambda}(1 - mq_2^*) > 0.$$

Therefore, the eigenvalues of A are strictly negative, and the dynamical system is locally exponentially stable. □

The theorem holds by invoking Theorem 1 in Ying (2016).

A.3 Proof of Lemma 3

The proof for this lemma proceeds along the lines of Theorem 1 in Ying (2016). Note that System 2 is also a CTMC. Let $W_m \in \{0, 1, 2\}$ denotes the state of the nth device. The state of the CTMC can be represented by the following vector:

$$X := \begin{bmatrix} \tilde{Q}_0 \\ \tilde{Q}_1 \\ \tilde{Q}_2 \\ Z_0 \\ Z_1 \\ Z_2 \end{bmatrix},$$

where

$$\tilde{Q}_i := \frac{1}{M} \sum_{m=1}^{M-1} 1_{W_m = i}$$

is the number of devices of the first $M-1$ devices in state i, averaged over M, and

$$Z_i := \frac{1}{M} 1_{W_M = i}.$$

Clearly, $\tilde{Q}_0 + \tilde{Q}_1 + \tilde{Q}_2 = \frac{M-1}{M}$ and $Z_0 + Z_1 + Z_2 = \frac{1}{M}$. Let us now follow the steps used in Ying (2016).

Let e_i denote a vector whose ith element is 1 and the rest are 0, then the transition rate of the CTMC is

$$\tilde{R}_{x,y} = \begin{cases} (M-1)\lambda\tilde{q}_0, \text{ if } y = x - \frac{1}{M}(e_1 - e_2) \\ \\ (M-1)d^*(1-(\tilde{q}_2+z_2))\tilde{q}_1, \text{if } y = x - \frac{1}{M}(e_2 - e_3) \\ \\ (M-1)\frac{1}{1+\lambda}\tilde{q}_2, \text{if } y = x - \frac{1}{M}(e_3 - e_1) \\ \\ \lambda z_0, \text{ if } y = x - \frac{1}{M}(e_4 - e_5) \\ \\ \alpha(x)(1-(\tilde{q}_2+z_2))z_1, \text{ if } y = x - \frac{1}{M}(e_5 - e_6) \\ \\ \frac{1}{1+\lambda}z_2, \text{ if } y = x - \frac{1}{M}(e_6 - e_1) \end{cases},$$

where $\alpha(x)$ is the sampling rate used by player M when the system is in state x.

We then define $f(x)$ to be

$$f(x) := \lim_{M\to\infty} \sum_{y:x\neq y} \tilde{R}_{x,y}(y-x),$$

and we have the following mean-field model

$$\begin{pmatrix} \dot{\tilde{q}}_0 \\ \dot{\tilde{q}}_1 \\ \dot{\tilde{q}}_2 \\ \dot{z}_1 \\ \dot{z}_1 \\ \dot{z}_2 \end{pmatrix} = f(x) = \begin{pmatrix} -\lambda\tilde{q}_0 + \dfrac{1}{1+\lambda}\tilde{q}_2 \\ \lambda\tilde{q}_0 - d^*(1-m\tilde{q}_2)\tilde{q}_1 \\ d^*(1-m\tilde{q}_2)\tilde{q}_1 - \dfrac{1}{1+\lambda}\tilde{q}_2 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{A.9}$$

which can be simplified to

$$\begin{pmatrix} \dot{\tilde{q}}_0 \\ \dot{\tilde{q}}_1 \\ \dot{\tilde{q}}_2 \end{pmatrix} = f(\tilde{q}) = \begin{pmatrix} -\lambda\tilde{q}_0 + \dfrac{1}{1+\lambda}\tilde{q}_2 \\ \lambda\tilde{q}_0 - d^*(1-m\tilde{q}_2)\tilde{q}_1 \\ d^*(1-m\tilde{q}_2)\tilde{q}_1 - \dfrac{1}{1+\lambda}\tilde{q}_2 \end{pmatrix} \tag{A.10}$$

and is identical to the mean-field model (2.4).

We already proved that (2.4) is locally exponentially stable and globally asymptotically stable. Next we bound $E\left[\|\tilde{Q}(\infty) - q^*\|^2\right]$, where q^* is the mean-field equilibrium point.

Let $g(\tilde{q})$ be the solution to the Poisson equation Ying (2016):

$$\nabla g(\tilde{q}) f(\tilde{q}) = \|\tilde{q} - q^*\|^2, \quad \forall \tilde{q}.$$

Following the analysis of Ying (2016), we have the following equation: (equation (8) in Ying (2016)):

$$E\left[\|\tilde{Q}(\infty) - q^*\|^2 + \|Z(\infty)\|^2\right] \tag{A.11}$$

$$= E\left[\nabla g(\tilde{Q}(\infty)) \cdot \left(f(\tilde{Q}(\infty)) - \sum_{y:y\neq X(\infty)} \tilde{R}_{X(\infty),y}((y) - X(\infty))\right)\right. \tag{A.12}$$

$$- \sum_{y:y\neq X(\infty)} \left(\tilde{R}_{X(\infty),y}\left(g(q_y) - g(\tilde{Q}(\infty))\right)\right. \tag{A.13}$$

$$\left.\left. - \tilde{R}_{X(\infty),y}\left(\nabla g(\tilde{Q}(\infty)) \cdot (q_y - \tilde{Q}(\infty))\right)\right)\right], \tag{A.14}$$

where $X(\infty) = (\tilde{Q}(\infty), Z(\infty))$ and $q_y = (y_1, y_2, y_3)$ (i.e. q_y includes the first three elements of y).

To apply the main result in Ying (2016), we note that the following conditions are valid:

- **Bounded state Transition-rate** There exists a constant c such that:

$$\frac{1}{M} E\left[\sum_{y:y\neq x} \tilde{R}_{x,y}\right] \leq c$$

- **Bounded state Transition** There exists a constant c_1 such that for any (x, y) with $\tilde{Q}_{x,y} > 0$:

$$\|y - x\| \leq \frac{c_1}{M}$$

- **Partial Derivative** $f(x)$ is clearly twice differentiable.

- **Stability** As stated above, $f(x)$ is globally asymptotically stable and locally exponentially stable.

Therefore, only the **Perfect Mean-field Model** condition is not satisfied. Without the perfect mean-field model condition, the following bound still holds:

$$E\Big[\sum_{y:y\neq X(\infty)} \tilde{R}_{X(\infty),y}\big(g(q_y) - g(\tilde{Q}(\infty))$$

$$- \nabla g(\tilde{Q}(\infty)) \cdot (q_y - \tilde{Q}(\infty))\big)\Big] \leq \frac{C_1}{M}.$$

Therefore, we only need to bound (A.12).

Since $\|\nabla g(q)\|$ is bounded by a constant according to the analysis in Ying (2016) when the system satisfies the stability condition. We now focus on the following term:

$$f(q) - \sum_{y:y\neq(x,z)} \tilde{R}_{x,y}(q_y - q),$$

and have

$$\begin{aligned}
&\left\| f(q) - \sum_{y:y\neq(x,z)} \tilde{R}_{x,y}(q_y - q) \right\| \\
&= \left\| (\lambda(1 - \frac{M-1}{M})(e_1 - e_2)) + d^*(1 - (x_2))(1 - \frac{M-1}{M}) \times \right., \\
&(e_2 - e_3) + \frac{1}{1+\lambda}(1 - \frac{M-1}{M})(e_3 - e_1) - d^* z_2 \frac{M-1}{M} \times \\
&\left.(e_2 - e_3) \right\| < \frac{2}{M}\|2\lambda + \frac{2}{\lambda} + d^*\|.
\end{aligned}$$

which concludes the proof.

APPENDIX B

PROOFS OF AGE DEPENDENT MAC

Note, from the definition of u_i, for any $\alpha \in [0, A]$ we have,

$$
\begin{aligned}
u_i \geq & \frac{\alpha + \lambda + 1/\delta}{A + \lambda + 1/\delta}\Bigg(\frac{(1-\gamma)\alpha}{1/\delta + \lambda + \alpha}\big(R_i + \beta u_{-2}\big) - c(\alpha) + \\
& \beta \frac{\gamma\alpha}{1/\delta + \lambda + \alpha} u_i + \beta\big(\frac{1/\delta}{1/\delta + \lambda + \alpha} u_{i+1}\big) \\
& + \beta\big(\frac{\lambda}{1/\delta + \lambda + \alpha} u_0\big)\Bigg)
\end{aligned}
$$

We can rewrite this inequality as,

$$
\begin{aligned}
u_i \geq & \frac{1}{A + \lambda + 1/\delta - \alpha\beta\gamma} \times \\
& \Big((1-\gamma)\alpha\big(R_i + \beta u_{-2}\big) - c(\alpha)(\alpha + \lambda + 1/\delta) \\
& + \beta \frac{1}{\delta} u_{i+1} + \beta\lambda u_0 \Big)
\end{aligned}
$$

for any α in $[0, A]$. But this implies,

$$
\begin{aligned}
u_i \geq & \sup_{\alpha \in [0,A]} \frac{1}{A + \lambda + 1/\delta - \alpha\beta\gamma} \times \\
& \Big((1-\gamma)\alpha\big(R_i + \beta u_{-2}\big) - c(\alpha)(\alpha + \lambda + 1/\delta) \\
& + \beta \frac{1}{\delta} u_{i+1} + \beta\lambda u_0 \Big)
\end{aligned}
$$

Now since, $[0, A]$ is compact, we know there exists some $\alpha \in [0, A]$ such that for any $\epsilon > 0$,

$$
\begin{aligned}
u_i \leq & \frac{\alpha + \lambda + 1/\delta}{A + \lambda + 1/\delta}\Bigg(\frac{(1-\gamma)\alpha}{1/\delta + \lambda + \alpha}\big(R_i + \beta u_{-2}\big) - c(\alpha) + \\
& \beta \frac{\gamma\alpha}{1/\delta + \lambda + \alpha} u_i + \beta\big(\frac{1/\delta}{1/\delta + \lambda + \alpha} u_{i+1}\big) \\
& + \beta\big(\frac{\lambda}{1/\delta + \lambda + \alpha} u_0\big)\Bigg) + \epsilon
\end{aligned}
$$

But this means,

$$
\begin{aligned}
u_i \leq & \frac{1}{A + \lambda + 1/\delta - \alpha\beta\gamma} \times \\
& \Big((1-\gamma)\alpha\big(R_i + \beta u_{-2}\big) - c(\alpha)(\alpha + \lambda + 1/\delta) \\
& + \beta \frac{1}{\delta} u_{i+1} + \beta\lambda u_0 \Big) + K\epsilon
\end{aligned}
$$

for an appropriate constant K. This justifies our claim.

B.2 Convergence to the Mean-Field Limit

Lemma 9. *The dynamical system described by (2) is globally asymptotically stable.*

Proof. We will use the following Lyapunov function to prove the result,

$$V := \sum_{i=0}^{K-1} |\epsilon_i| + |\epsilon_K| + |\epsilon_{-1}| + |\epsilon_{-2}| \tag{3}$$

Global asymptotic stability is ensured when the Lyapunov function has a negative drift but is positive for all points except 0.
We split the function into more easy to manipulate parts, let V_1 be the first two terms of V, let $V_1 := \sum_{i=(1,0)}^{1,K-1} |\epsilon_i| + |\epsilon_K|$ and let $V_2 := V_1 + |\epsilon_{-2}|$.

We will now look at the derivative of V_1.

$$\frac{dV_1}{dt} = \sum_{\epsilon_i>0,i=(0)}^{K-1} \frac{d\epsilon_i}{dt} - \sum_{\epsilon_i<0,i=(0)}^{K-1} \frac{d\epsilon_i}{dt} + \frac{d|\epsilon_K|}{dt}$$

Collecting all the terms with ϵ_j, $K > j > 0$ together, assuming ϵ_{j+1} and ϵ_j are both positive we get:

$$\epsilon_j\big(- (\lambda + \frac{1}{\delta} + \alpha_j(1-\gamma) + \frac{1}{\delta})\big) = -\epsilon_j\big(\lambda + \alpha_j(1-\gamma)\big)$$

if they are both negative :

$$\epsilon_j\big(\lambda + \alpha_j(1-\gamma)\big)$$

if they have different signs:

$$-|\epsilon_j|\big(\lambda + 2\frac{1}{\delta} + \alpha_j(1-\gamma)\big)$$

Therefore, the terms containing $\epsilon_j \leq -|\epsilon_j|\big(\lambda + \alpha_j(1-\gamma)\big)$ and the term containing $\epsilon_K < -\lambda|\epsilon_K|$. The term containing $\epsilon_{-1} \leq -|\epsilon_{-1}|\big(\lambda + \frac{1}{\delta} + \alpha_0(1-\gamma)\big)$. This leads us to the following :

$$\begin{aligned}\frac{dV_1}{dt} < &- \sum_{j=(0)}^{K-1} |\epsilon_j|\big(\lambda + \alpha_j(1-\gamma)\big) - \lambda|\epsilon_K| \\ &+ m\epsilon_2\Bigg(\sum_{\epsilon_i>0,j=0}^{j=K-1} \pi_j\alpha_j - \sum_{\epsilon_i<0,j=0}^{j=K-1} \pi_j\alpha_j\Bigg) \\ &- \lambda\epsilon_{-2}\end{aligned}$$

When $\epsilon_0 > 0$ and

$$\begin{aligned}\frac{dV_1}{dt} &< -\sum_{j=(0)}^{K-1} |\epsilon_j|\big(\lambda + \alpha_j(1-\gamma)\big) - \lambda|\epsilon_K| \\ &\quad + m\epsilon_2\Bigg(\sum_{\epsilon_i>0, j=0}^{j=K-1} \pi_j\alpha_j - \sum_{\epsilon_i<0, j=0}^{j=K-1} \pi_j\alpha_j\Bigg) \\ &\quad + \lambda\epsilon_{-2} \\ &\text{when } \epsilon_0 < 0\end{aligned} \tag{4}$$

ottherwise.
We are now ready to consider $\frac{dV_2}{dt}$. Recall that :

$$\frac{d\epsilon_{-2}}{dt} = \sum_{j=0}^{K-1} \epsilon_j\alpha_j(1-\gamma) - m\epsilon_{-2}\sum_{j=0}^{K-1} \pi_j\alpha_j - \frac{1}{1+\lambda}\epsilon_{-2}$$

Now if $\epsilon_{-2} > 0$ then

$$\frac{d|\epsilon_{-2}|}{dt} = \sum_{j=0}^{K-1} \epsilon_j\alpha_j(1-\gamma) - m|\epsilon_{-2}|\sum_{j=0}^{K-1} \pi_j\alpha_j - \frac{1}{1+\lambda}\epsilon_{-2}$$

and if $\epsilon_{-2} < 0$ then

$$\frac{d|\epsilon_{-2}|}{dt} = -\sum_{j=0}^{K-1} \epsilon_j\alpha_j(1-\gamma) + m\epsilon_{-2}\sum_{j=0}^{K-1} \pi_j\alpha_j + \frac{1}{1+\lambda}\epsilon_{-2}$$

which leads us to,

$$\frac{d|\epsilon_{-2}|}{dt} \leq \sum_{j=0}^{K-1} |\epsilon_j|\alpha_j(1-\gamma) - m|\epsilon_{-2}|\sum_{j=0}^{K-1} \pi_j\alpha_j - \frac{1}{1+\lambda}|\epsilon_{-2}|$$

Plugging this inequality in $V_2 := V_1 + |\epsilon_{-2}|$ to compute the drift, we get:

$$\begin{aligned}\frac{dV_2}{dt} &:= \frac{dV_1}{dt} + \frac{d|\epsilon_{-2}|}{dt} \\ &< -\sum_{j=(1,0)}^{K-1} |\epsilon_j|\big(\lambda + \alpha_j(1-\gamma)\big) - \lambda|\epsilon_K|\end{aligned}$$

$$+m\epsilon_2\Bigg(\sum_{\epsilon_i>0, j=0}^{j=K-1} \pi_j\alpha_j - \sum_{\epsilon_i<0, j=0}^{j=K-1} \pi_j\alpha_j\Bigg) - \lambda|\epsilon_{-2}|$$

$$
\begin{aligned}
&+\sum_{j=0}^{K-1}|\epsilon_j|\alpha_j(1-\gamma)-m|\epsilon_{-2}|\sum_{j=0}^{K-1}\pi_j\alpha_j-\frac{1}{1+\lambda}|\epsilon_{-2}| \\
\leq&-\lambda\sum_{j=(1,0)}^{1,K-1}|\epsilon_{1,j}|+|\epsilon_2|(\lambda-\frac{1}{\lambda+1})-\lambda|\epsilon_K|
\end{aligned}
$$

Finally, we can now bound the derivative for V:

$$
\begin{aligned}
\frac{dV}{dt}:=&\frac{dV_2}{dt}+\frac{d|\epsilon_{-1}|}{dt} \\
&<-\lambda\sum_{j=(1,0)}^{1,K-1}|\epsilon_j|+|\epsilon_{-2}|(\lambda-\frac{1}{\lambda+1})-\lambda|\epsilon_{-1}| \\
&+|\epsilon_{-2}|\frac{1}{(\lambda+1)}-\lambda|\epsilon_K| \\
&\leq-\lambda\sum_{j=(0)}^{K-1}|\epsilon_j|+|\epsilon_{-2}|\lambda-\lambda|\epsilon_{-1}|-\lambda|\epsilon_K|
\end{aligned}
$$

Further, by definition $\sum_{j=(0)}^{K-1}\epsilon_j+\epsilon_{-1}+\epsilon_{-2}+\epsilon_K=0$. So, $\sum_{j=(0)}^{K-1}\epsilon_j+\epsilon_{-1}+\epsilon_K=-\epsilon_2$. From the triangle inequality we have, $|\epsilon_2|\leq\sum_{j=(0)}^{K-1}|\epsilon_j|+|\epsilon_{-1}|+|\epsilon_K|$ which gives us:

$$
\frac{dV}{dt}<\lambda(|\epsilon_2|-\sum_{j=(0)}^{K-1}|\epsilon_j|-|\epsilon_{-1}|-|\epsilon_K|)\leq 0
$$

□

We may now proceed to show (in the next lemma) that the system is in fact globally exponentially stable.

Lemma 10. *We can use equation (3) to show that the system is globally exponentially stable.*

Proof. This proof is simply a more careful handling of equation (3) where instead of looking for the simplest upper bounds to derive the results we take a more systematic case by case approach. Consider the Lyapunov function (3). We proceed by finding the derivative as before and using equation (4) for the function V_1. We break the system into the following cases:
Case I: $\big(\epsilon_{-1}<0,\epsilon_0>0,\epsilon_{-2}<0\big)$

Returning to the equation,

$$\begin{aligned}
\frac{dV_1}{dt} <& -\sum_{j=(0)}^{K-1} |\epsilon_j|\big(\lambda + \alpha_j(1 - m\pi_{-2})\big) - \lambda|\epsilon_K| \\
& + m\epsilon_{-2}\Bigg(\sum_{\epsilon_i>0,j=0}^{j=K-1} \pi_j\alpha_j - \sum_{\epsilon_i<0,j=0}^{j=K-1} \pi_j\alpha_j\Bigg) \\
& - \lambda\epsilon_2 \\
=& -\sum_{j=(0)}^{K-1} |\epsilon_j|(\lambda + \alpha_0(1 - m\pi_{-2})) + m\epsilon_{-2}\pi_0\alpha_0 \\
& + m\epsilon_{-2}\Bigg(\sum_{\epsilon_i>0,j=1}^{j=K-1} \pi_j\alpha_j - \sum_{\epsilon_i<0,j=1}^{j=K-1} \pi_j\alpha_j\Bigg) \\
& - \lambda\epsilon_{-2} - \lambda|\epsilon_K|
\end{aligned}$$

This allows us to bound the derivative of V_2.

$$\begin{aligned}
& \frac{dV_2}{dt} < \\
& -\sum_{j=(1,0)}^{K-1} |\epsilon_j|\big(\lambda + \alpha_j(1 - \gamma)\big) - \lambda|\epsilon_K| + m\epsilon_{-2}\pi_0\alpha_0 \\
& + m\epsilon_{-2}\Bigg(\sum_{\epsilon_j>0,j=1}^{j=K-1} \pi_j\alpha_j - \sum_{\epsilon_j<0,j=1}^{j=K-1} \pi_j\alpha_j\Bigg) - \lambda\epsilon_{-2} \\
& + \sum_{j=0}^{K-1} \epsilon_j\alpha_j(1-\gamma) - m|\epsilon_{-2}|\sum_{j=0}^{K-1} \pi_j\alpha_j - \frac{1}{1+\lambda}|\epsilon_{-2}| \\
& - \lambda|\epsilon_K| \\
\leq & -\lambda\sum_{j=(1,0)}^{K-1} |\epsilon_j| + |\epsilon_{-2}|(\lambda - \frac{1}{\lambda+1} - 2m\pi_0\alpha_0) - \lambda|\epsilon_K|
\end{aligned}$$

Now we can find a bound on $\frac{dV}{dt}$:

$$\begin{aligned}
\frac{dV}{dt} :=& \frac{dV_2}{dt} + \frac{d|\epsilon_{-1}|}{dt} \\
<& -\lambda \sum_{j=(0)}^{K} |\epsilon_j| + |\epsilon_{-2}|(\lambda - \frac{1}{\lambda+1} - 2m\pi_0\alpha_0) - \lambda|\epsilon_K| \\
& - \lambda|\epsilon_{-1}| + |\epsilon_{-2}|\frac{1}{(\lambda+1)} \\
\leq& -\lambda \sum_{j=(0)}^{K-1} |\epsilon_j| + |\epsilon_{-2}|(\lambda - 2m\pi_0\alpha_0) - \lambda|\epsilon_{-1}| - \lambda|\epsilon_K| \\
& - \lambda|\epsilon_{-1}| + |\epsilon_{-2}|\frac{1}{(\lambda+1)} - \lambda|\epsilon_K| \\
=& -\lambda \sum_{j=0}^{K} |\epsilon_j| + |\epsilon_{-2}|(\lambda - 2m\pi_0\alpha_0)) - \lambda|\epsilon_{-1}| - \lambda|\epsilon_K|
\end{aligned}$$

If $(\lambda - 2m\pi_0\alpha_0) < 0$ then let ρ_I be the minimum of λ and $2m\pi_0\alpha_0 - \lambda$, then $\frac{dV}{dt} < -\rho_I V$ otherwise, if $(\lambda - 2m\pi_0\alpha_0) \geq 0$ then

$$\begin{aligned}
\frac{dV}{dt} =& -(\lambda - m\pi_0\alpha_0)\left(\sum_{j=(1,0)}^{K-1} |\epsilon_j| + |\epsilon_{-1}| - |\epsilon_{-2}| + |\epsilon_K|\right) \\
& - m\pi_0\alpha_0 V \\
\leq& - m\pi_0\alpha_0 V
\end{aligned}$$

Where the second inequality follows from $\epsilon_{-2} = \sum_{j=0}^{K} \epsilon_j + \epsilon_{-1}$ and the triangle inequality.
So $\rho_I = \alpha_0 m\pi_{(0)}$ and $\frac{dV}{dt} < -\rho_I V$.
By symmetry we can make the same conclusion when $\big(\epsilon_{-1} > 0, \epsilon_0 < 0, \epsilon_2 > 0\big)$. Next we consider the case ,

Case II: $\epsilon_0 > 0, \epsilon_{-2} > 0$

This time we can bound the derivative of V_1 by,

$$\begin{aligned}
\frac{dV_1}{dt} <& - \sum_{j=(1,0)}^{K-1} |\epsilon_j|\big(\lambda + \alpha_j(1-\gamma)\big) - \lambda|\epsilon_K| \\
& + m\epsilon_{-2}\left(\sum_{\epsilon_j>0, j=0}^{j=K-1} \pi_j\alpha_j - \sum_{\epsilon_j<0, j=0}^{j=K-1} \pi_j\alpha_j\right) - \lambda|\epsilon_{-2}|
\end{aligned}$$

$$\le -\lambda \sum_{j=(1,0)}^{K-1} |\epsilon_j| - \lambda|\epsilon_K| - |\epsilon_2|(\lambda - \frac{1}{\lambda+1})$$
$$+ m\epsilon_{-2}\Big(\sum_{\epsilon_j>0,j=0}^{j=K-1} \pi_j\alpha_j - \sum_{\epsilon_j<0,j=0}^{j=K-1} \pi_j\alpha_j\Big)$$

Next bounding the derivative for V_2 :

$$\frac{dV_2}{dt} < -\sum_{j=0)}^{K-1} |\epsilon_j|\big(\lambda + \alpha_j(1-\gamma)\big) - \lambda|\epsilon_K|$$
$$+ m\epsilon_{-2}\Big(\sum_{\epsilon_j>0,j=0}^{j=K-1} \pi_j\alpha_j - \sum_{\epsilon_j<0,j=0}^{j=K-1} \pi_j\alpha_j\Big) - \lambda|\epsilon_{-2}|$$
$$+ \sum_{j=0}^{K-1} \epsilon_j\alpha_j(1-\gamma) - m|\epsilon_{-2}|\sum_{j=0}^{K-1} \pi_j\alpha_j - \frac{1}{1+\lambda}|\epsilon_{-2}|$$
$$\le -\lambda \sum_{j=(1,0)}^{K-1} |\epsilon_j| - |\epsilon_{-2}|(\lambda - \frac{1}{\lambda+1}) - \lambda|\epsilon_K|$$

It follows that $\frac{dV}{dt} < -\lambda V$. The same can be shown when $\epsilon_0 < 0, \epsilon_2 < 0$. The last case we shall consider is,

Case III: $\big(\epsilon_{-1} > 0, \epsilon_0 > 0, \epsilon_2 < 0\big)$ The derivative for the Lyapunov function is:

$$\frac{dV}{dt} := \frac{dV_2}{dt} + \frac{d|\epsilon_{-1}|}{dt}$$
$$< -\lambda \sum_{j=(1,0)}^{1,K} |\epsilon_{1,j}| + |\epsilon_2|(\lambda - \frac{1}{\lambda+1}) - \lambda|\epsilon_{-1}|$$
$$- |\epsilon_2|\frac{1}{(\lambda+1)}$$
$$\le -\lambda \sum_{j=(0)}^{K-1} |\epsilon_j| + |\epsilon_2|(\lambda - \frac{2}{\lambda+1}) - \lambda|\epsilon_{-1}|$$
$$= \lambda \sum_{j=(0)}^{K} |\epsilon_j| + |\epsilon_2|(\lambda - \frac{2}{\lambda+1}) - \lambda|\epsilon_0|$$

If $(\lambda - \frac{2}{\lambda+1} < 0$ then let ρ_I be the minimum of λ and $\frac{2}{\lambda+1} - \lambda)$, then $\frac{dV}{dt} < -\rho_{III}V$ otherwise, if $(\lambda - \frac{2}{\lambda+1}) \ge 0$ then

$$\frac{dV}{dt} = -(\lambda - m\pi_{(0)}\alpha_0)\Big(\sum_{j=(0)}^{K-1} |\epsilon_j| + |\epsilon_{-1}| - |\epsilon_{-2}| + |\epsilon_K|\Big)$$
$$- m\pi_{(0)}\alpha_0 V$$
$$\le -m\pi_{(0)}\alpha_0 V$$

So $\rho_{III} = \frac{1}{\lambda+1}$ and $\frac{dV}{dt} < -\rho_{III}V$. The same can be shown when

$$\left(\epsilon_{-1} < 0, \epsilon_0 < 0, \epsilon_2 > 0\right)$$

.

Therefore, by setting ρ to be the minimum of λ, ρ_I and ρ_{III} we can ensure that the drift of the lyapunov function is given by $\frac{dV}{dt} = -\rho V$ □

B.2.1 Proof of lemma (4)

Proof. Our proof of this lemma will be similar to the proof of Proposition (1). Recall that α_i^* is the optimal policy for state i which ranges from $\{0, 1, 2....\}$. As in Proposition (1), we use $u_i(\alpha)$ to indicate that at state i a device uses the policy α while fixing u_{-i}^*, that is, the valuation function at each other state remains fixed at its optimal value. Then, by definition $u_i^* := u_i(\alpha_i^*)$.From Proposition (1) we know,

$$u_i^* > u_{i+1}^*$$

Now consider $u_i(\alpha_{i+1}^*)$. that is the optimal policy for state $i+1$ used in state i, clearly,

$$u_i^* - u_{i+1}^* \geq u_i(\alpha_{i+1}^*) - u_{i+1}^*$$

. Similarly, we can upper bound $u_i^* - u_{i+1}^*$ by $u_i^* - u_{i+1}(\alpha_i^*)$. This gives us,

$$u_i^* - u_{i+1}(\alpha_i^*) \geq u_i^* > u_{i+1}^* \geq u_i(\alpha_{i+1}^*) - u_{i+1}^*$$

Substituting the expressions above, we get:

$$\frac{(R_i - R_{i+1})(1-\gamma)\alpha_i^* + \beta 1/\delta(u_{i+1}^* - u_{i+2}^*)}{A - \alpha_i\beta\gamma + \lambda + 1/\delta} \geq$$
$$\frac{(R_i - R_{i+1})(1-\gamma)\alpha_{i+1}^* + \beta 1/\delta(u_{i+1}^* - u_{i+2}^*)}{A - \alpha_{i+1}\beta\gamma + \lambda + 1/\delta}$$

Note that,

$$h(x) := \frac{(R_i - R_{i+1})(1-\gamma)x + \beta 1/\delta(u_{i+1}^* - u_{i+2}^*)}{A - x\beta\gamma + \lambda + 1/\delta}$$

is a monotonically increasing function in x while $A - x\beta\gamma + \lambda + 1/\delta$ is greater than zero. Therefore,

$$h(\alpha_i^*) \geq h(\alpha_{i+1}^*) \iff \alpha_i^* \geq \alpha_{i+1}^*$$

□

B.2.2 Proof of Proposition (2)

Proof. Fix $\tilde{\epsilon} > 0$. Let $u_i^* = u_i(\alpha^*)$ as in the notation above and let $\tilde{u}_i = u_i(\alpha^K)$. We will begin by showing $u_0^* - \tilde{u}_0 < C_u\tilde{\epsilon}$ for some constant C_u for sufficiently large K

when $\beta < \beta_0$.

$$
\begin{aligned}
u_0^* - \tilde{u}_0 =& \Bigg(\frac{\beta\lambda}{\lambda + 1/\delta + \alpha_0(1-\gamma\beta)} (u_0^* - \tilde{u}_0) \\
& + \eta \frac{\alpha_0(1-\gamma)}{\lambda + 1/\delta + \alpha_0(1-\gamma\beta)} (u_0^* - \tilde{u}_0) \Bigg) \\
& + \frac{\beta 1/\delta}{\lambda + 1/\delta + \alpha_0(1-\gamma\beta)} (u_1^* - \tilde{u}_1^*) \\
& < \beta \Bigg(\frac{\lambda}{\lambda + 1/\delta} + \eta \Bigg) (u_0^* - \tilde{u}_{-1}) \\
& + \frac{\beta\lambda}{\lambda + 1/\delta} \Bigg(\beta \Bigg(\frac{\lambda}{\lambda + 1/\delta} + \eta \Bigg) (u_0^* - \tilde{u}_0) \\
& + \frac{\beta\lambda}{\lambda + 1/\delta} (u_2^* - \tilde{u}_2) \Bigg)
\end{aligned}
$$

Expanding $u_j^* - \tilde{u}_j$ for $j > 0$ to $j < K$ we get:

$$
\begin{aligned}
u_0^* - \tilde{u}_0 <& \beta \Bigg(\frac{\lambda}{\lambda + 1/\delta} + \eta \Bigg) (u_0^* - \tilde{u}_0) \sum_{i=0}^{K} \Big(\frac{\beta 1/\delta}{\lambda + 1/\delta} \Big)^i \\
& + \Big(\frac{\beta 1/\delta}{\lambda + 1/\delta} \Big)^K (u_{K+1}^* - \tilde{u}_K) \\
<& \beta \Bigg(\frac{\lambda}{\lambda + 1/\delta} + \eta \Bigg) (u_0^* - \tilde{u}_0) \sum_{i=0}^{\infty} \Big(\frac{\beta 1/\delta}{\lambda + 1/\delta} \Big)^i \\
& + \Big(\frac{\beta 1/\delta}{\lambda + 1/\delta} \Big)^K (u_{K+1}^* - \tilde{u}_{K+1})
\end{aligned}
$$

For sufficiently large K, we can ensure that $\big(\frac{\beta 1/\delta}{\lambda + 1/\delta}\big)^{K/2} \frac{R}{1-\beta} < \tilde{\epsilon}$.

$$
\begin{aligned}
u_0^* - \tilde{u}_0 <& \beta \Bigg(\frac{\lambda}{\lambda + 1/\delta} + \eta \Bigg) \frac{1}{1 - \beta \frac{1/\delta}{1/\delta + \lambda}} (u_0^* - \tilde{u}_0) + \tilde{\epsilon} \\
<& \beta \Bigg(1 + \eta \frac{\lambda + 1/\delta}{\lambda} \Bigg) (u_0^* - \tilde{u}_0) + \tilde{\epsilon}
\end{aligned}
$$

If we fix β_0 to be such that $\beta_0 \big(1 + \eta \frac{\lambda + 1/\delta}{\lambda}\big) < 1$, then $(u_0^* - \tilde{u}_0) < C_u \tilde{\epsilon}$. We can use the same reasoning to show that $u_i^* - \tilde{u}_i < C_u \tilde{\epsilon}$ for $0 < i < K/2$. We have shown that for sufficiently large K we can ensure that $|u_i^* - \tilde{u}_i| < C_u \tilde{\epsilon}$ for $0 \leq i < K/2$.

For any policy $x := \{x_0, x_1, x_{-2} .....\}$ with fixed γ, one may use the detail balance equations derived from the Markov chain in Fig (1) to find the fraction of time π_j a device spends in state j given by:

$$\pi_0 = \frac{\lambda(1-\pi_{-2})}{\lambda + 1/\delta + x_0(1-\gamma)}$$

and

$$\pi_j = \pi_{j-1}\frac{1/\delta}{\lambda + 1/\delta + x_j(1-\gamma)}$$

for $j > 0$.
and

$$\pi_{-1} = \frac{1}{\lambda(1+\lambda)}\pi_{-2}$$

This gives us the following equation for the fraction of time spent in state -2 by a device,

$$1 = \pi_{-2}(1 + \frac{1}{\lambda(1+\lambda)}) + (1-\pi_{-2})\frac{\lambda}{1/\delta + \lambda + x_0(1-\gamma)} \times \sum_{i=0}^{\infty} \Pi_{j=1}^{i} \frac{1/\delta}{1/\delta + \lambda + x_j(1-\gamma)}.$$

.

Now set,

$$\kappa := 1 + \frac{1}{\lambda(1+\lambda)} \tag{B.1}$$

and

$$\theta(x) := \frac{\lambda}{1/\delta + \lambda + x_0(1-\gamma)} \sum_{i=0}^{\infty} \Pi_{j=1}^{i} \frac{1/\delta}{1/\delta + \lambda + x_j(1-\gamma)} \tag{B.2}$$

and the fraction of time a device spends in state -2 is given by :

$$\pi_{-2}(x) = \frac{1-\theta(x)}{\kappa - \theta(x)} \tag{B.3}$$

. Clearly, there exists K large enough so that $|\pi_{-2}(\alpha) - \pi_{-2}(\alpha^{(K)})| < \epsilon$. since π_i converges geometrically to zero we can ensure that,

- $K\epsilon \to 0$ as K tends to infinity.

- $\sum_{i=K/2}^{\infty} \pi_i < \epsilon$ for both α^* and α^K.

- $|\pi_i(\alpha^*) - \pi_i(\alpha^K)| < \epsilon$ for all $0 < i < K/2$.

- As above $|u_i^* - \tilde{u}_i| < \epsilon$ for all $0 < i < K/2$.

Consider, $\quad \mathrm{E}_X\{u(X,\alpha^*)\} - E_{\tilde{X}}\{u(\tilde{X},\alpha^{(K)})\}$

$= |\sum_{i=1}^{K/2}\left(\pi_i(\alpha)u_i^* - pi_i(\alpha^K)\tilde{u}_i\right) + \pi_0(\alpha^*)u_0^* - \pi_0(\alpha^K)\tilde{u}_0$

$+ \pi_{-2}(\alpha^*)u_{-2} - \pi_{-2}(\alpha^K)\tilde{u}_{-2} + \sum_{i=K/2+1}^{\infty}\left(\pi_i(\alpha^*)u_i^* - \pi_i(\alpha^K)\tilde{u}_i\right)|$

Bounding $u_i^*(x)$ by $\frac{R}{1-\beta}$, replacing

$$\pi_i(\alpha^*)u_i^* - \pi_i(\alpha^K)\tilde{u}_i^*$$

with

$$\pi_i(\alpha^K)(u_i^* - \tilde{u}_i) + u_i^*(\pi_i(\alpha^*) - \pi_i(\alpha^K))$$

and using the triangle inequality we get,

$$\begin{aligned}
&E_X\{u(X,\alpha^*)\} - E_{\tilde{X}}\{u(\tilde{X},\alpha^{(K)})\} \\
&< \sum_{=0}^{K/2}|\pi_i(\alpha^*) - \pi_i(\alpha^K)|\tilde{u_i^*} + \sum_{=0}^{K/2}|u_i^* - \tilde{u_i}|\pi_i(\alpha^K) \\
&+ |\pi_{-2}(\alpha^*) - \pi_{-2}(\alpha^K)|u_{-2} + |\pi_0(\alpha) - \pi_0(\alpha^K)|u_0 \\
&+ |u_0^* - \tilde{u_0}|\pi_0(\alpha^K) + |u_{-2}^* - \tilde{u_{-2}}|\pi_{-2}(\alpha^K) + R\frac{\epsilon}{1-\beta} \\
&< \epsilon\left(\frac{R}{1-\beta}(\frac{K}{2}+2)+1\right)
\end{aligned}$$

for sufficiently large K we then have for any $\hat{\epsilon} > 0$,

$$E_X\{u(X,\alpha^*)\} - E_{\tilde{X}}\{u(\tilde{X},\alpha^{(K)})\} < \hat{\epsilon}$$

□

B.3 Existence of LMFNE

Proving T_1 is Lipschitz in α under the L1-norm and hence in the ∞-norm turns out to be much simpler than showing T_2 is Lipschitz in γ. Therefore, we start by showing that T_1 is Lipschitz in α in the next lemma while proving T_2 is Lipschitz over the two lemmas following that.

Lemma 11. *For T_1 and policy α as defined in Section (3.2), T_1 is Lipschitz in $\|\alpha\|_\infty$.*

Proof. From equation (B.1), (B.2) and (B.3) in section 3.3. The fraction of time spent for a device to be in state (-2) is given by:

$$\pi_{-2} = \frac{1-\theta(\alpha)}{\kappa-\theta(\alpha)}$$

We now examine $\|\frac{\partial \pi_{-2}}{\partial \alpha_i}\|$ for all i

$$\left|\frac{\partial \pi_{-2}}{\partial \alpha_i}\right| = \left|\frac{\kappa+1-2\theta(\alpha)}{(\kappa-\theta(\alpha))^2}\right|\left|\frac{\partial \theta(\alpha)}{\partial \alpha_i}\right|.$$

We can explicitly compute $\left|\frac{\partial \theta(\alpha)}{\partial \alpha_i}\right|$ as follows:

$$\left|\frac{\partial \theta(\alpha)}{\partial \alpha_0}\right| = \frac{(1-\gamma)}{\lambda + 1/\delta + \alpha_0(1-\gamma)} \frac{\lambda}{1/\delta + \lambda + \alpha_0(1-\gamma)} \times \sum_{k=0}^{\infty} \Pi_{j=1}^{k} \frac{1/\delta}{1/\delta + \lambda + \alpha_j(1-\gamma)}$$

and

$$\left|\frac{\partial \theta(\alpha)}{\partial \alpha_i}\right| = \frac{(1-\gamma)}{\lambda + 1/\delta + \alpha_i(1-\gamma)} \frac{\lambda}{1/\delta + \lambda + \alpha_0(1-\gamma)} \Pi_{n=1}^{i} \frac{1/\delta}{1/\delta + \lambda + \alpha_n(1-\gamma)} \sum_{k=0}^{\infty} \Pi_{j=1}^{k} \frac{1/\delta}{1/\delta + \lambda + \alpha_{i+j}(1-\gamma)}. \quad \text{(B.4)}$$

Next we will bound $\left|\frac{\partial \theta(\alpha)}{\partial \alpha_i}\right|$ simply by replacing every α_i with 0 in the expression above to obtain :

$$\left|\frac{\partial \theta(\alpha)}{\partial \alpha_0}\right| \leq \frac{(1-\gamma)}{\lambda + 1/\delta} \frac{\lambda}{1/\delta + \lambda} \sum_{k=0}^{\infty} \left(\frac{1/\delta}{1/\delta + \lambda}\right)^k = \frac{(1-\gamma)}{\lambda + 1/\delta}$$

. Similarly, we have

$$\left|\frac{\partial \theta(\alpha)}{\partial \alpha_i}\right| \leq \frac{(1-\gamma)}{\lambda + 1/\delta} \left(\frac{1/\delta}{1/\delta + \lambda}\right)^i.$$

Finally, summing the bounds we get:

$$\sum_{i=0}^{\infty} \left|\frac{\partial \pi_2}{\partial \alpha_i}\right| \leq \left|\frac{\kappa + 1 - 2\theta(\alpha)}{(\kappa - \theta(\alpha))^2}\right| \frac{(1-\gamma)}{\lambda + 1/\delta} \sum_{i=0}^{\infty} \left(\frac{1/\delta}{1/\delta + \lambda}\right)^i$$

The gradient is therefore bounded. π_2 is Lipschitz in α under the l_1 norm and hence, Lipschitz in the l_∞ norm. □

Next we show u_i is Lipschitz in γ for fixed α.

Lemma 12. *If $\beta < \beta_0$, then u_i is Lipschitz in γ for all i.*

Proof. Consider, $u_i(\alpha_i^*(\gamma), \gamma)$, where $\alpha_i^*(\gamma)$ is the optimal policy at state i given the fraction of busy channels γ. Following the notation in Proposition (1), $u_i(\alpha_i, \hat{\gamma})$ indicates the value of the utility for state i given a fraction of busy channels $\hat{\gamma}$ when the rate is equal to α_i. Concretely, if $\{u_i^*(\hat{\gamma})\}_i$ is the optimal value function given $\hat{\gamma}$ then,

$$\begin{aligned} u_i(\alpha_i, \hat{\gamma}) =& \frac{1}{1/\delta + \lambda + A - \hat{\gamma}\beta\alpha_i} \Big\{ (1-\hat{\gamma})\alpha_i\big(R_i + \beta u_{-2}^*(\hat{\gamma})\big) \\ & -c(\alpha_i)\left(\alpha_i + \lambda + 1/\delta\right) + \frac{\beta}{\delta} u_{i+1}^*(\hat{\gamma}) + \beta\lambda u_0^*(\hat{\gamma}) \Big\} \end{aligned}$$

u_i^* is both an implicit and explicit function of γ, computing how $\alpha_i^*(\gamma)$ varies with γ would immediately tell us how T_2 varies with γ. However, to do so we need to first compute also compute how u_i^* varies with γ. We will avoid this problem as follows; Let us define $\triangle_{i,\gamma_1,\gamma_2}$ by:

$$\triangle_{i,\gamma_1,\gamma_2} := |u_i(\alpha_i^*(\gamma_1), \gamma_1) - u_i(\alpha_i^*(\gamma_2), \gamma_2)|$$

For fixed (γ_1, γ_2), $(u_{1,i}(\alpha^*(\gamma_1), \gamma_1), u_i(\alpha^*(\gamma_2), \gamma_2))$ are two positive real numbers. Without loss of generality assume $u_i(\alpha^*(\gamma_1), \gamma_1) > u_i(\alpha^*(\gamma_2), \gamma_2)$. Then,

$$\begin{aligned} \triangle_{i,\gamma_1,\gamma_2} =& |u_i(\alpha_i^*(\gamma_1), \gamma_1) - u_i(\alpha_i^*(\gamma_2), \gamma_2)| \\ \leq& |u_i(\alpha_i^*(\gamma_1), \gamma_1) - u_i(\alpha_i^*(\gamma_1), \gamma_2)| \end{aligned}$$

For ease of notation in the rest of the section, let

$$D := A + \lambda + 1/\delta,$$

let

$$L(\alpha, \gamma) := \frac{\beta}{\delta} u_{i+1}^* + \beta\lambda u_0^* - c(\alpha)(\alpha + \lambda + 1/\delta)$$

and let $\alpha = \alpha^*(\gamma_1)$.
Now,

$$u_i(\alpha, \gamma_1) = \frac{1}{D - \alpha\beta\gamma_1}\bigg(\alpha(1-\gamma_1)(R_i + \beta u_{-2}^*(\gamma_1)) + L(\alpha, \gamma_1)\bigg)$$

The first term can be rewritten as,

$$\begin{aligned} (1-\gamma_1)\alpha(R_i + \beta u_{-2}^*(\gamma_1)) =& (1-\gamma_2)\alpha(R_i + \beta u_{-2}^*(\gamma_1)) + R_i(\gamma_2 - \gamma_1)\alpha \\ =& (1-\gamma_2)\alpha(R_i + \beta u_{-2}^*(\gamma_2) + \beta\triangle_{-2,\gamma_1,\gamma_2}) + R_i(\gamma_2 - \gamma_1)\alpha \\ =& (1-\gamma_2)\alpha(R_i + \beta u_{-2}^*(\gamma_2) + \eta\beta^2\triangle_{0,\gamma_1,\gamma_2}) + R_i\alpha(\gamma_2 - \gamma_1) \end{aligned}$$

and the second term, $L(\alpha, \gamma_1)$ can be rewritten as,

$$L(\alpha, \gamma_1) = L(\alpha, \gamma_2) + \frac{\beta}{\delta}\triangle_{i+1,\gamma_1,\gamma_2} + \beta\lambda\triangle_{0,\gamma_1,\gamma_2}$$

Substituting into

$$u_i(\alpha(\gamma_1), \gamma_1) - u_i(\alpha(\gamma_1), \gamma_2)$$

we get,

$$
\begin{aligned}
&\frac{1}{D-\alpha\beta\gamma_1}R_i(\gamma_2-\gamma_1)\alpha+\\
&\bigg(\alpha(1-\gamma_2)(R_i+u^*_{-2}(\gamma_2))+L(\alpha,\gamma_2)\bigg)\times\\
&\left[\frac{1}{D-\alpha\gamma_1\beta}-\frac{1}{D-\alpha\gamma_2\beta}\right]+\frac{1}{D-\alpha\gamma_1\beta}\times\\
&\bigg[(1-\gamma_1)\alpha\beta^2\eta\triangle_{0,\gamma_1,\gamma_2}+\beta 1/\delta\triangle_{i+1,\gamma_1,\gamma_2}+\beta\lambda\triangle_{0,\gamma_1,\gamma_2}\bigg]
\end{aligned}
$$

Collecting the terms with $\gamma_2-\gamma_1$, we get:

$$
(\gamma_2-\gamma_1)\left[\frac{1}{D-\alpha\beta\gamma_1}R_i\alpha+\left(\frac{\alpha\beta}{(D-\alpha\gamma_1\beta)(D-\alpha\gamma_2\beta)}\right)\times\right.
$$

$$
\left.\bigg(\alpha(1-\gamma_2)(R_i+u^*_{-2}(\gamma_2))+L(\alpha,\gamma_2)\bigg)\right]
$$

Note that $\frac{1}{D-\alpha\gamma_2\beta}(1-\gamma_2)\alpha(R_i+u_{-2}(\gamma_2))+L(\alpha,\gamma_2)$ is bounded above by $\frac{R_0}{1-\beta}$. This gives us the following bound,

$$
\begin{aligned}
&(\gamma_2-\gamma_1)\left[\frac{1}{D-\alpha\beta\gamma_1}R_i\alpha+\left(\frac{\alpha\beta}{(D-\alpha\gamma_1\beta)(D-\alpha\gamma_2\beta)}\right)\times\right.\\
&\left.\bigg(\alpha(1-\gamma_2)(R_i+u_{-2}(\gamma_2))+L(\alpha,\gamma_2)\bigg)\right]
\end{aligned}
$$

$$
\begin{aligned}
&\le(\gamma_2-\gamma_1)\frac{\alpha}{D-\alpha\gamma_1\beta}\left(R_0+\frac{\beta R_0}{1-\beta}\right)\\
&=(\gamma_2-\gamma_1)\frac{\alpha}{D-\alpha\gamma_1\beta}\frac{R_0}{1-\beta}
\end{aligned}
$$

We can bound $\frac{\alpha}{D-\alpha\gamma_1\beta}\frac{R_0}{1-\beta}$ by a constant C_1 that is independent of i or γ. Similarly let $\phi_0=\frac{1}{D-\alpha\gamma_1\beta}\bigg((1-\gamma_1)\alpha\beta^2\eta+\beta\lambda\bigg)$ and let $\phi_1=\frac{\beta 1/\delta}{D-\alpha\gamma_1\beta}$. Note, both ϕ_0 and ϕ_1 are less than 1, we now have:

$$
\triangle_{i,\gamma_1,\gamma_2}\le C_1|\gamma_1-\gamma_2|+\triangle_{0,\gamma_1,\gamma_2}\phi_0+\triangle_{i+1,\gamma_1,\gamma_2}\phi_1
$$

Expanding $\triangle_{i+1,\gamma_1,\gamma_2}$ in the same way as we expanded $\triangle_{i,\gamma_1,\gamma_2}$ we get,

$$
\begin{aligned}
\triangle_{i,\gamma_1,\gamma_2}\le C_1|\gamma_1-\gamma_2|+\phi_0\triangle_{0,\gamma_1,\gamma_2}+\phi_1C_1|\gamma_1-\gamma_2|\\
+\phi_0\phi_1\triangle_{0,\gamma_1,\gamma_2}+\phi_1^2\triangle_{i+2,\gamma_1,\gamma_2}
\end{aligned}
$$

This leads us to the following bound for any $\triangle_{i,\gamma_1,\gamma_2}$:

$$\triangle_{i,\gamma_1,\gamma_2} \leq \frac{C_1}{1-\phi_1}|\gamma_1 - \gamma_2| + \frac{\phi_0}{1-\phi_1}\triangle_{0,\gamma_1,\gamma_2}$$

In particular this bound holds for $\triangle_{0,\gamma_1,\gamma_2}$:

$$\triangle_{0,\gamma_1,\gamma_2} \leq \frac{C_1}{1-\phi_1}|\gamma_1 - \gamma_2| + \frac{\phi_0}{1-\phi_1}\triangle_{0,\gamma_1,\gamma_2}$$

with a little shuffling we get:

$$\triangle_{0,\gamma_1,\gamma_2} \leq \frac{C_1}{1-\phi_0-\phi_1}|\gamma_1 - \gamma_2|$$

Note that for $\beta < \beta_0$, $\phi_0 + \phi_1 < 1$ for all γ_1 and γ_2. Thus, for any γ_1 and γ_2, $\frac{\triangle_{0,\gamma_1,\gamma_2}}{\gamma_2-\gamma_1}$ is bounded. $\square$

We will use this result to prove that T_2 is Lipschitz in γ. Consider some fixed γ_1, let $\alpha^{(1)} := \{\alpha_i^{(1)}\}_i$ and $\alpha^{(2)} := \{\alpha_i^{(2)}\}_i$ be the optimal policy when the fraction of busy channels are γ_1 and γ_2 respectively. Additionally, let $\alpha^{(j)}$, with $j = \{1, 2\}$ be in $\mathcal{P}^K$. If γ_1, γ_2 are such that, $|\gamma_1 - \gamma_2| < \epsilon$ then, from the lemma above, $\sup_i |u_i(\gamma_1, \alpha_i^{(1)}) - u_i(\gamma_2, \alpha_i^{(2)})| < C_1\epsilon$. We are now ready to prove that T_2 is continuous in γ.

Lemma 13. *The map T_2 is continuous in γ.*

Proof. Using the convention from lemma 12 we have,

$$u_i(\alpha_i^{(1)}, \gamma_1) = \frac{1}{D - \alpha_i^{(1)}\beta\gamma_1}\bigg(\alpha_i^{(1)}(1-\gamma_1)(R_i + \beta u^*_{-2}(\gamma_1)) \\ + L(\alpha_i^{(1)}, \gamma_1)\bigg)$$

Similarly,

$$u_i(\alpha_i^{(2)}, \gamma_2) = \frac{1}{D - \alpha_i^{(2)}\beta\gamma_2}\bigg(\alpha_i^{(2)}(1-\gamma_2)(R_i + \beta u^*_{-2}(\gamma_2)) \\ + L(\alpha_i^{(2)}, \gamma_2)\bigg)$$

Using the first order Taylor expansion of $\frac{1}{D-\alpha_i^{(2)}\beta\gamma_2}$ about γ_1 for sufficiently small ϵ we get,

$$u_i(\alpha_i^{(2)}, \gamma_2) = \left[\frac{1}{D - \alpha_i^{(2)}\beta\gamma_1} + \frac{\epsilon\alpha_i^{(2)}\beta}{(D - \alpha_i^{(2)}\beta\gamma_1)^2}\right] \times \left(\alpha_i^{(2)}(1-\gamma_2)(R_i + \beta u_{-2}^*(\gamma_2)) + L(\alpha_i^{(2)}, \gamma_2)\right)$$

From lemma 12,

$$L(\alpha_i^{(2)}, \gamma_2) < L(\alpha_i^{(1)}, \gamma_1) + C_1\epsilon(\beta\lambda + \beta 1/\delta) + c(\alpha_i^{(1)})(\alpha_i^{(1)} + \lambda + 1/\delta) - c(\alpha_i^{(2)})(\alpha_i^{(2)} + \lambda + 1/\delta)$$

similarly,

$$\alpha_i^{(2)}(1-\gamma_2)(R_i + \beta u_{-2}^*(\gamma_2)) <$$

$$\alpha_i^{(1)}(1-\gamma_1)(R_i + \beta u_{-2}^*(\gamma_2)) + A\epsilon(R_i + \beta u_{-2}^*(\gamma_2)) + (\alpha_i^{(2)} - \alpha_i^{(1)})(1-\gamma_2)(R_i + \beta u_{-2}^*(\gamma_2))$$

We can use these upper bounds to bound $u_i(\alpha_i^{(2)}, \gamma_2)$.

$$u_i^*(\alpha_i^{(2)}, \gamma_2) < \left[\frac{1}{D - \alpha_i^{(2)}\beta\gamma_1} + \frac{\epsilon\alpha_i^{(2)}\beta}{(D - \alpha_i^{(2)}\beta\gamma_1)^2}\right] \times$$

$$\Big(\alpha_i^{(1)}(1-\gamma_1)(R_i + \beta u_{-2}^*(\gamma_1)) + L(\alpha_i^{(1)}, \gamma_1) +$$

$$A\epsilon(R_i + \beta u_{-2}^*(\gamma_2)) + C_1\epsilon(\alpha_i^{(1)}(1-\gamma_1)\beta + \beta\lambda + \beta 1/\delta) +$$

$$(\alpha_i^{(2)} - \alpha_i^{(1)})(1-\gamma_2)(R_i + \beta u_{-2}^*(\gamma_2) +$$

$$c(\alpha_i^{(1)})(\alpha_i^{(1)} + \lambda + 1/\delta) - c(\alpha_i^{(2)})(\alpha_i^{(2)} + \lambda + 1/\delta)\Big)$$

Note that each term that is multiplied with ϵ is bounded. We will use C_2 for a bound on the sum of these terms. This gives us,

$$u_i(\alpha_i^{(2)}, \gamma_2) < \left[\frac{1}{D - \alpha_i^{(2)}\beta\gamma_1}\right] \times$$

$$\Big(\alpha_i^{(1)}(1-\gamma_1)(R_i + \beta u_{-2}(\gamma_1)) + L(\alpha_i^{(1)}, \gamma_1) +$$

$$(\alpha_i^{(2)} - \alpha_i^{(1)})(1-\gamma_2)(R_i + \beta u_{-2}^*(\gamma_2) +$$

$$c(\alpha_i^{(1)})(\alpha_i^{(1)} + \lambda + 1/\delta) - c(\alpha_i^{(2)})(\alpha_i^{(2)} + \lambda + 1/\delta)\Big) + C_2\epsilon$$

From lemma 12,

$$|u_i(\alpha_i^{(1)}, \gamma_1) - u_i(\alpha_i^{(2)}, \gamma_1)| < C_1\epsilon$$

Substituting the upper bound obtained into the inequality above we get,

$$\begin{aligned}
&\Bigg|\left[\frac{1}{D - \alpha_i^{(1)}\beta\gamma_1} - \frac{1}{D - \alpha_i^{(2)}\beta\gamma_1}\right]\times \\
&\left(\alpha_i^{(1)}(1-\gamma_1)(R_i + \beta u_{-2}(\gamma_1)) + L(\alpha_i^{(1)}, \gamma_1)\right)+ \\
&\frac{1}{D - \alpha_i^{(1)}\beta\gamma_1}\Bigg((\alpha_i^{(2)} - \alpha_i^{(1)})(1-\gamma_2)(R_i + \beta u_{-2}(\gamma_2)+ \\
&c(\alpha_i^{(1)})(\alpha_i^{(1)} + \lambda + 1/\delta) - c(\alpha_i^{(2)})(\alpha_i^{(2)} + \lambda + 1/\delta)\Bigg)\Bigg| \\
&< (C_1 + C_2)\epsilon
\end{aligned}$$

It follows that $\sup_i |\alpha_i^{(1)} - \alpha_i^{(2)}| \xrightarrow{\gamma^{(1)} \to \gamma^{(2)}} 0$. From Proposition (2), there are only finitely many non zero α_i, this means that $\sum_{i=0}^{\infty} |\alpha_i^{(1)} - \alpha_i^{(2)}| \leq K \sup_i |\alpha_i^{(1)} - \alpha_i^{(2)}|$ which converges to zero as $\gamma^{(1)}$ converges to $\gamma^{(2)}$. Thus, α is continuous in γ under the L1 norm. Therefore, the map T_2 is continuous in γ. □

B.3.1 Proof of Theorem (9)

Proof. Let us begin by fixing $K \geq K_1$. For any policy, $x = \{x_0, x_1, ...x_{K-1}, 0, 0, ...\} \in \mathcal{P}^K$, by Theorem (1), the system converges to the mean field limit, given by:

$$\pi_0(x) = \frac{\lambda}{\lambda + 1/\delta + x_0(1-\gamma)}$$

$$\pi_j(x) = \pi_{j-1}(x)\frac{1/\delta}{\lambda + 1/\delta + x_j(1-\gamma)}$$

for $K > j > 0$, and

$$\pi_K(x) = \frac{1}{\lambda\delta}\pi_{K-1}(x)$$

Clearly, $\pi_i(x)$ is geometrically decreasing in i upto $K-1$ with rate smaller than $\frac{1/\delta}{\lambda+1/\delta}$. Hence, for any $\eta > 0$, there exists a finite $N(\eta)$ independent of K (order, $\mathcal{O}(log(1/\eta))$ such that

$$\sum_{j=-2}^{N(\eta)} \pi_j(x) > 1 - \eta$$

Since, this is true for any $x \in \mathcal{P}^K$ and $\mathcal{P}^K \subset \mathcal{P}^{K+1} \subset \mathcal{P}^{K+2}...$, we know that $\{\pi(x)|x \in \mathcal{P}^K\}$ is tight (there exists a compact set, S, such that the measure of S is greater than or equal to $1-\epsilon$ i,e, , $\pi^K(x)(S) = \sum_{j \in S} \pi_j(x) > 1-\epsilon$ for all $x \in \mathcal{P}^K$). Since our state space is a separable metric space under the topology of point-wise convergence, by Prokhorov's theorem, $\{\pi(x)|x \in \mathcal{P}^K\}$ is relatively compact and metrizable. Similarly, note that the sequence of MFNE, $\{\pi^{K_i}\}$ is relatively compact. Therefore, the sequence $\{\pi^{K_i}\}$, contains a convergent subsequence, $\pi^{K_{i_j}} \to \pi^*$. Defining a sequence, $j(i) = i_j$ completes the first part of the theorem.

The map $T_{\text{proj}} : \pi \to \gamma$ is continuous, since $\gamma = \pi_{-2}$ for any distribution π (it is simply the projection map). Now, T_2 is continuous in γ hence, $T_2 \circ T_{\text{proj}}$ is continuous in π. The sequence $\{T_2 \circ T_{\text{proj}}(\pi^{K_i})\}$ must also be relatively compact. Since, T_2 is continuous and maps from $[0,1] \to l_\infty$ under the supremum norm, it must be continuous in the topology of pointwise convergence. Therefore, the sub-sequence $T_2 \circ T_{\text{proj}}(\pi^{K_{j_i}})$ is convergent and converges to $\alpha = \{\alpha_0, \alpha_1 ...\}$. The fixed point for the system of ODEs is simply the map $T_1(\alpha)$ which is also convergent under the weak topology. $\square$

APPENDIX C

INTRODUCTION TO MEAN FIELD GAME

C.1 A general Model

Suppose we start with a set of players P, for the sake of an initial model. We will start with a finite set, $\#(P) = N$. Each player can choose an action α^i from an action set A. The agents inhabit a state space X, for example in a congestion game where the agents have to travel to a location, the state space maybe their position. To be more concrete, we assume our state space is $X \subset \mathbb{R}^d$ for some finite d. For the rest of the paper we will assume that an agents state evolves as a Brownian motion which is dependent on the action taken by the player as well as the state. Mathematically, if a player i takes action α_t^i we state that a given players state changes based on the stochastic differential equation,

$$dX_t^i = b(t, X_t^i, \alpha_t^i, X_t^{-i})dt + \sigma(X_t^i, \alpha_t^i)dW_t^i \tag{C.1}$$

Where W_t^i is a $k \times 1$ dimensional Wiener process. A small side note, the action set is the set of all controls for $t \geq 0$ and is W_t^i adapted. Each player, i has a cost function J_i that he wishes to minimize, if we wish to consider a congestion game, then the cost function can be a model for fuel efficiency.
Suppose the N agents fix their actions to be $(\alpha^1, \alpha^2, ...\alpha^N)$, let α^i denote the action chosen by player i and let a_{-i} denote the actions taken by all the other agents, additionally assume that the initial position of the agents is fixed and given. Then, according to Lasry and Lions (2007) we can define J_i from $A^N \to \mathbb{R}$ by,

$$J^i(\alpha^i, \alpha^2...\alpha^N) = \inf_{\alpha^i} E\left[\int_0^T L^i(X_t^i, \alpha_t^i) + F^i(X_t^1, ...X_t^N)dt\right] \tag{C.2}$$

We assume that $L^i(X_T^i, \alpha_T^i) + F^i(X_T^1, ...X_T^N)$ are given final values for the equation above. Here, we assume that F^i is Lipschitz on X^N, L^i is Lipschitz in $x^i \in X$ and uniformly in α^i bounded, finally,

$$\inf_{x^i} L^i(x^i, \alpha^i)/|\alpha^i| \to \infty \tag{C.3}$$

when $|\alpha| \to \infty$. For a full exposition on the necessity of these conditions to find the optimal control, the reader is directed to chapter 3, Evans (2010) section (3.3.2) onward or Chapter 4, Yong and Zhou (2010). A Nash point α^* here is defined in the conventional sense, for fixed α_{-i},

$$J^i(\alpha^i, \alpha_{-i}) \geq J^i(\alpha^*, \alpha_{-i}) \tag{C.4}$$

Remark 7.

- Here we will force σ in (C.1) to lie in $\mathcal{L}^2([0,T], \mathbb{R}^{d\times k})$ (under some appropriate norms) and we would like σ to always remain positive definite. The idea is to use the Burkholder-Davis-Gundy Inequality inequality to show uniqueness of (C.1) for fixed α^i.

- Here, if $m^i(t)$ denotes the law of the random variable X_t^i then, at first glance it would appear that the evolution of $m^i(t)$ will be very hard to track. The reason for this specific form of (C.1) is under certain conditions we will be able to assume asymptotic independence of the agents state distribution, that is, if m^N is the joint distribution of the agents then, under suitable conditions we are able to replace the joint distribution with the product distribution, $\Pi_j m_t^j$. In statistical physics this phenomenon is called ***Propagation of chaos***. The most famous text on the topic is Sznitman's book, Sznitman (1991).

It is clear here that both (C.1) and (C.2) will depend on the joint distribution of all the agents, m^N. We would like to examine the asymptotic regime, $N \to \infty$ to approximate the behaviour of individual agents in order to simplify the problem. In the regime where $N \to \infty$ if we can show $m^N \to m$ for some appropriate limiting distribution m we may have some hope to jointly solve the (C.1) and (C.2). It turns out that this is too much to ask under the most general conditions, the joint evolution of the system will follow some variant of the fokker plank equation. We will now try to simplify (C.2).
Let,

$$m := \frac{1}{N-1} \sum_{j \neq i} \delta_{x_j} \tag{C.5}$$

denote the empirical distribution of the agents. Then, we will replace F^i by an operator $V[m](x)$ from the space of probability measures on X into a bounded set of Lipschitz functions on X,

$$V[\frac{1}{N-1} \sum_{j \neq i} \delta_{x_j}](x^i) = F^i(x^1, x^2 ... x^N) \tag{C.6}$$

A good example of a class of such functions $V[m](x) := F(K \star m(x), x)$ where K is any Lipschitz function on $\mathbb{R}^d \times X \to \mathbb{R}$ and $K \star m(x) := \int_X K(x,y)m(y)dy$. Additionally we will assume that the operator V is continuous in m, that is, $V[m_n]$ converges to $V[m]$ whenever, m_n converges to m.

Remark 8.
Note, whenever X is compact, the set of empirical measures is always tight (in the sense of Prokhorov) and therefore, the set of empirical distributions is always relatively compact. Thus, so long as V is continuous, it maps a relatively compact set to a relatively compact set. This fact will come particularly handy when we want to talk about the existence of convergent sub-sequences!

Further, we will perform the same treatment on the motion of the agents,

$$dX_t^i = b(t, X^i, \alpha, m)dt + \sigma(t, X^i, \alpha^i)dW_t \tag{C.7}$$

It turns out that given the states evolve according to the Brownian motion described above we cannot have a more general setup. For example, we cannot really allow σ to depend on any other state except X_t^i and α_t^i. Essentially, this allows us to interpret states evolving according to the optimal control plus some state, action dependent noise given by a Wiener process W_t^i.

We will now summarize the model we have come up with so far. Since all the cost functions are now the same, the agents are homogeneous. We can then compute the solution of the problem described so far by computing the solution for one representative agent.

Representative agent's problem

For fixed empirical distribution m, solve:

$$J(a(.), m) = \inf_{a \in A} E\left[\int L^i(X_t^i, \alpha_t^i) + V[m](x) \right] \tag{C.8}$$

This equation will be solved backwards since, at time T we are given the terminal conditions for the control problem. Subject to,

$$dX_t^i := b(t, X^i, \alpha^i, m)dt + \sigma(t, X^i, \alpha^i)dW_t^i \tag{C.9}$$

We are given initial conditions for this SDE which we can use to deduce the trajectory of any agent. Finally given the initial positions of all the agents it is now possible to compute the empirical distribution of the system. Thus, given the initial conditions we can solve in the forward direction the combined distribution of the agents m.

Remark 9.

- Readers familiar with the Stochastic Maximum principle will notice that the notion of solving the system forward and backward is commonly referred to as the forward-backward stochastic differential equation.

- Instead of considering $N-1$ individual agents the representative agent now computes the combined effect of the other players through the empirical measure. In the limit as N tends to infinity this measure may be easier to model using either ODEs or PDEs. The modelling of the limiting behaviour in this way is referred to as our mean field approximation.

- In the remark above we talked about the agent approximating the combined effect of the other agents using the limiting behaviour as $N \to \infty$. However, technically the model we will be using is in fact going to assume the continuum limit of agents rather than the more intuitive countable limit. It turns out that in scenarios where we can use propagation of chaos, technically either forms of infinite might be used although this is far from the scope of this report. Interested readers may check Carmona (2004).

A mean field game equilibrium is a consistency condition (α^*, m^*) such that given distribution m^* an agent chooses control α^* and given all agents choose α^* the combined density function of the agents is given by m^*

C.2 Hamilton Jacobi Bellman Equation

Let the rest of the actions of the agents be fixed. Additionally, the agent believes that he might not be able to influence the empirical distribution m. This problem boils down to a single agent trying to optimize their path in space time with respect to a fixed cost function.
We are going to use the calculus of variation for the deterministic simple case to introduce (at least heuristically) the method the agent may use to solve their problem (most of the derivation for this section and the next can be found both in Evans (2010) and TT (2010)). For now let us simplify the problem as follows, let the agent choose his velocity vector at every instance in time, this is his policy α. We will also simplify the motion of the agent, he now simply moves as $dx = \alpha dt$,the cost function is now:

$$J(x_0) = \inf_{\alpha} \int_0^T L(\alpha)dt + J(x_T)$$

where x_T is some arbitrary end point and x_0 is a given initial position at time 0. Now we can generalize the initial point and the initial time, instead of starting at 0 we now start at some arbitrary time t_0 between 0 and T. Let, (with a lot of abuse of notation)

$$J(x_0, t_0) = \inf_{\alpha} \int_{t_0}^T L(\alpha)dt + J(x_T)$$

Presumably, given that the agent starts at (x_0, t_0) and solves the problem correctly at this time, he moves to a new location, $x_0 + \alpha dt$, therefore, the agents problem now is to solve the problem at $(x_0 + \alpha dt, t_0 + dt)$, combining the two we have,

$$J(x_0, t_0) = \inf_{\alpha} J(x_0 + \alpha dt, t_0 + dt) + L(\alpha)dt$$

using first order Taylor expansion,

$$J(x_0, t_0) = \inf_{\alpha} \left[J(x_0, t_0) + dt\big[\frac{\partial J(x_0, t_0)}{\partial t} + \alpha.\nabla J(x_0, t_0) + L(\alpha)\big] \right]$$

we are choosing α to minimize the expression above, essentially, choosing it such that we minimize the term multiplied by dt should yield the minimizer. Now if we further assume the convexity of L we should be able to obtain a unique minimizer,

$$\alpha^* = \inf_{\alpha} \big[\frac{\partial J(x_0, t_0)}{\partial t} + \alpha.\nabla J(x_0, t_0) + L(\alpha)\big]$$

Define the *Hamiltonian* for L by,

$$H(p) := \sup_{\alpha} \alpha.p - L(\alpha)$$

then, $-H(\nabla J(t_0, x_0))$ is going to minimize $\alpha.\nabla J(x_0, t_0) + L(\alpha)$, therefore, we can equivalently write the problem as, find α such that,

$$-\frac{\partial J(x_0, t_0)}{\partial t} + H(\nabla J(t_0, x_0)) = 0$$

however, our choice of (x_0, t_0) are arbitrary, this will lead us to the Hamilton Jacobi Bellman equation,

$$-\frac{\partial J}{\partial t} + H(\nabla J) = 0 \tag{C.10}$$

This equation is being solved back in time starting from T onward. From the method of characteristics for this PDE we obtain the following ODE,

$$\frac{dx}{dt}(s) = -\frac{\partial H}{\partial p}(p(s), x(s)) = \alpha(s) \tag{C.11}$$

$$\frac{dp}{dt}(s) = \frac{\partial H}{\partial x}(p(s), x(s)) \tag{C.12}$$

While the equations derived above are a useful tool in optimal control, we need to export this process to a stochastic setting. Suppose for the problem above, additionally, the motion is now, $dx = \alpha dt + \sigma dW_t$. The cost function remains the same, except we will need to take an expectation over the trajectory of the agent,

$$J(x_0) = \inf_{\alpha} E\left[\int_0^T L(\alpha)dt + J(x_T)\right]$$

using the same treatment as before, we have, the change in position for a small amount of time is now $(x_0 + \alpha dt + \sigma dW_t, t_0 + dt)$:

$$J(x_0, t_0) = \inf_{\alpha} EJ(x_0 + \alpha dt + \sigma dW_t, t_0 + dt) + L(\alpha)dt$$

writing out first order terms in t and second order terms for the Wiener process, one may use *Ito's formula* to obtain (due to quadratic variation of the brownian motion):

$$\begin{aligned} J(t_0, x_0) =& E\Bigg[J(x_0, t_0) + \frac{\partial J(x_0, t_0)}{\partial t}dt + \alpha.\nabla J(x_0, t_0)dt + \sigma dW_t.\nabla J(x_0, t_0) \\ &+ \frac{\sigma^2}{2}\nabla^2 J(x_0, t_0)dW_t^2 + L(\alpha)dt\Bigg] \end{aligned}$$

Now, we are given x_0, t_0 and we have inherently assumed that our policy is $\mathcal{F}_t$ adapted, so we can compute the expectation, this means the W_t term will go to 0 and the W_t^2 term will become t,

$$J(t_0, x_0) = J(x_0, t_0) + \left[\frac{\partial J(x_0, t_0)}{\partial t} + \alpha.\nabla J(x_0, t_0) + \frac{\sigma^2}{2}\nabla^2 J(x_0, t_0) + L(\alpha)\right]dt$$

Substituting our definition of the Hamiltonian and replacing $\sigma^2/2$ by ν we now obtain what is called the *Viscous Hamilton Jacobi Bellman* equations,

$$-\frac{\partial J}{\partial t} - \nu\Delta J + H(\nabla J) = 0 \tag{C.13}$$

C.3 Fokker Plank Equation

As in the previous section, we are going to begin with a simple model and move up toward a more complete model for our setting. It should be noted here that we are inherently assuming that the agents are identical in cost function and in the functions b and σ and in our setting we already have a propagation of chaos result available. It should also be noted that our derivation will be an informal sketch rather than a complete formal proof.
Rather than deal with each of the agents separately, we will pass to a continuum limit $N \to \infty$ and consider just the (normalised) density function $m(t, x)$ of the agents, which is a non-negative function with total mass $\int_{\mathbb{R}^d} m(t, x)\ dx = 1$ for each time t. Informally, for an infinitesimal box in space $[x, x + dx]$, the number of agents in that box should be approximately $Nm(t, x)|dx|$.
Suppose the velocity at each point in space time, $\alpha(x, t)$ is given to us as well as the initial distribution of the agents $m(0)$. Let G be some convenient smooth test function, we are looking for ways to describe m as a weak solution (often called distributional solution in PDEs) concept to some PDE. We will show how this can be done using G.
Intuitively,

$$\int_{\mathbb{R}^d} m(t, x)G(x)\ dx \approx \frac{1}{N} \sum_{i=1}^{N} G(x_i(t))$$

Taking a time derivative and noting that G is explicitly independent of time we get,

$$\int_{\mathbb{R}^d} \frac{\partial}{\partial t}(m(t, x))G(x)dx \approx \frac{1}{N} \sum_{i=1}^{N} \alpha(t, x_i(t))\nabla G(x_i(t))$$

In the continuum limit, the right hand side can be rewritten as,

$$\frac{1}{N} \sum_{i=1}^{N} \alpha(t, x_i(t))G(x_i(t)) \to \int_{\mathbb{R}^d} \alpha(t, x)m(t, x)\nabla G(x)dx$$

integrating by parts we get,

$$-\int_{\mathbb{R}^d} \nabla.(\alpha(t, x)m(t, x))G(x)dx$$

Here we are ignoring the constant term from integration by parts since we are free to choose $G(x)$ to be any candidate function, in particular one whose terminal values are 0. combining the equations we now have,

$$\int_{\mathbb{R}^d} \left(\frac{\partial}{\partial t}(m(t, x)) + \nabla.(\alpha(t, x)m(t, x))\right)G(x)dx = 0$$

This gives us the following equation, (known as the advection equation)

$$\frac{\partial}{\partial t}(m(t, x)) + \nabla.(\alpha(t, x)m(t, x)) = 0$$

Using the procedure from the previous section to export this derivation to the stochastic setting, thanks to Ito's formula we get the *Fokker Plank equation* :

$$\frac{\partial}{\partial t}(m(t,x)) + \nabla.(\alpha(t,x)m(t,x)) - \nu\Delta m(t,x) = 0 \tag{C.14}$$

Note this equation is solved forward in time given our initial condition $m(0,x)$ for all x in the domain.

C.4 Main Results

In the previous section we presented the tools needed to solve our problem. We are now going to use these tools to present the main result that we have been building up to so far. We need some general conditions on the state and action space in order to show our results,

Assumptions for Compactness and smoothness assumptions

1. Assume that the state space $X \subset \mathbb{R}^d$ is compact.

2. Assume that the action set is A is compact.

3. We are given our agent moves according to (C.9) and has cost function (C.8)

4. With regards to the (C.9), assume that the initial distribution of the agents are smooth.

5. The cost function (C.8) is convex and at least C^1.

6. Let us define our Hamiltonian for by

$$H(x,p) = \sup_{\alpha}(p.\alpha - L(x,\alpha)) \tag{C.15}$$

7. (C.3) holds, additionally assume that

$$\exists\theta \in (0,1), \inf_{x}\left(\nabla H.p + \frac{\theta}{d}\nu^i(H)^2\right) > 0 \text{for } |p| \text{ large} \tag{C.16}$$

,

Theorem 14. *Let N agents optimize their costs according to* (C.9) *and has cost function* (C.8) *under the assumptions above, then, for fixed, finite N, Nash equilibria exist given by these 3 N tuples,* $(\lambda_1^N, \lambda_2^N...\lambda_N^N)$, $(J_1^N, ..J_N^N)$ *and* $(m_1^N, ..m_N^N)$ *which satisfy the following:*

- *The optimal control policy is given by:*

$$\alpha_i = -\frac{\partial}{\partial p} H(x, \nabla J_i) \tag{C.17}$$

- *Each λ_i^N is bounded in $\mathbb{R}$, J_i^N is relatively compact in $C^2(X)$, m_i^N is relatively compact in the space of probability measures of X. This is true for all $N \in \mathbb{N}$*

- $$\sup_{i,j} \left(|\lambda_i^N - \lambda_j^N| + \|J_i^N - J_j^N\|_\infty + \|m_i^N - m_j^N\|_\infty \right) \to 0$$

  as $N \to \infty$.

- *Any converging subsequence $(\lambda_i^{N_k}, J_i^{N_k}, m_i^{N_k})_{N_k}$ converges to some triplet (λ, J, m) satisfying*

$$-\nu \Delta J + H(x, \nabla J) + \lambda - V[m] = 0 \tag{C.18}$$

$$-\frac{\partial}{\partial t}(m(t,x)) + \nabla.\left(\frac{\partial}{\partial p} H(x, \nabla J) m(t,x) \right) + \nu \Delta m(t,x) = 0 \tag{C.19}$$

Remark 10.

- The theorems here can be split into two parts, the first part about the existence of a Nash equilibrium the second about convergence of the Nash equilibrium to the continuum limit which are given by the solutions of PDE (C.18) and (C.19).

- The existence theorem can actually be generalized a lot further, however, in the absence of the second part of the theorem, (the convergence to the solution of the PDE) the computation of the Nash equilibrium quickly becomes intractable.

- Underlying the main results of the theorem is an inherent result (propagation of chaos) that allows us to use the law of large numbers to compute, m using the different m_i^N. Infact, in any given problem of mean field games, one of the hardest parts to prove is the existence of a closed form limit for the joint distribution of the agents. Our choice of trajectory and cost function play a critical role in allowing us to leverage results that already exist in statistical physics. In general the evolution may follow what are called Mckean Vlasov equations which are not easy to solve. Dawson (1995)

Proof. We will now present a very rough sketch of the proof.

- Let us assume that we have a fixed distribution m for all time t and x.

- A typical agent i can now compute his own best response (optimal control) policy using

$$-\nu_i \Delta v_i + H(x, \nabla v_i) + \lambda_i = \int V[m](x) dm \tag{C.20}$$

- Now, given that each agent has computed their optimal policy, there is a fixed action set for all the other agents α_{-i}. The agent can now compute the emprical distribution over time, assuming the initial distribution $m_0(0, x)$ is known to him by computing

$$-\frac{\partial m_j}{\partial t} + \nu^j \Delta m_j + \nabla.\big(\alpha_j m_j\big) = 0 \tag{C.21}$$

- Using the law of large numbers agent i can now extrapolate to find the empirical distribution of the joint system of the agents $\hat{m}$.

- We have now defined a map from $\Phi : \mathcal{M}(X) \to \mathcal{M}(X)$, where $\mathcal{M}$ is the space of probability distribution on X. $\Phi(m) = \hat{m}$

- Due to assumption 2, using Prokhorov theorem, we know that the set of m is relatively compact.

- Due to assumption 1, 5, 6 and 7, the Hamilton Jacobi bellman equation says that the optimal policy for each agent i is unique, what is more the map from m to α_i is continuous. The joint action space A^N is compact. Therefore, the map $m \to (\alpha_1 ... \alpha_N)$ is continuous and maps to a relatively compact set.

- Finally, due to assumption 4 the Fokker Plank equations yield a continuous solution to our problem and the map from $(\alpha_1 ... \alpha_N) \to \hat{m}$ is continuous (the Hopf maximum principle for elliptical operators).

- Therefore, the map from $\Phi : m \to \hat{m}$ is continous, the set of distributions are relatively compact, which implies that the closure is compact and they are obviously convex. Hence, by Shauder's fixed point theorem, therefore, for each N a fixed point exists.

- Due to the relative compactness of the set of distributions, we now know, for the sequence m^N has a convergent subsequence N_k. The last set of results come from verifying that each agent i will satisfy the equations (C.18) and (C.19).

□