Analysis of Automotive Cyber-Attacks on Highways using Partial Differential Equation Models

Meysam Ghanavati[1], Animesh Chakravarthy[2], Prathyush P. Menon[3]

***Abstract*— This paper considers scenarios wherein a group of malicious vehicles on a highway perform a cooperative attack with the motive of creating undesirable wave effects among other vehicles on the highway. The two species of vehicles - malicious vehicles and normal vehicles, and their associated interaction effects, are modeled using Partial Differential Equations (PDEs). The malicious vehicles, which may be arbitrarily distributed on the highway, perform a sequence of velocity changes with the objective of making the density/velocity profile on the highway, track a reference profile. This reference profile (chosen by the malicious vehicles) has the property that once generated, it spontaneously evolves into a shock wave that propagates along the highway. Analytical expressions governing the velocity inputs of the malicious vehicles with which they can generate such waves are determined, for perfect as well as imperfect information scenarios. Simulation results are presented to validate the theory.**

I. Introduction

Rapid advancement of technology is transforming our cities into what are now referred to as "Smart-Cities", which are urban centers that integrate cyber-physical technologies and infrastructure, such as public transportation, energy, gas and water distribution, to enhance the quality of life of citizens. While ensuring better efficiency and convenience, the increased connectivity also expands the potential attack surface for malicious actors [1]. There are many scenarios wherein malicious cyber-attacks can occur. Examples include attacks on the smart grid [2],[3], gas transmission and distribution networks [4], large-scale process engineering plants [5], [6], water networks, UAVs, and automobiles [7]-[9]. As autonomous vehicles become prevalent in the connected transport infrastructure, it is possible that an attacker may try to hack the driving software of some of the vehicles, and thereby introduce undesirable effects on other vehicles on the highway [8]-[11].

This paper discusses one such scenario, wherein a group of malicious vehicles on a highway perform a cooperative attack with the motive of creating undesirable abrupt wave effects on other vehicles on the highway. More specifically, the malicious vehicles can perform a series of (subtle) velocity changes that will cause the density and velocity profiles of the non-malicious (normal) vehicles to attain certain pre-specified reference spatial distributions. These reference spatial velocity/density profiles can be so chosen by the malicious vehicles, that once formed, they subsequently evolve into undesirable effects such as shock waves, traffic jams, stop-and-go-traffic, etc. Shock waves are particularly undesirable because they can lead to multi-vehicle pile-up crashes.

This paper models and analyzes such attacks in a Partial Differential Equation (PDE) framework. The use of PDE models for studying traffic flows has a fairly long history. The earliest such model was the Lighthill-Whitham-Richards (LWR) model [12]. The LWR model is basically a first order hyperbolic PDE based on a gas dynamic-like continuity equation, which represents the conservation of cars on a highway. Subsequently, second order hyperbolic models have been developed, for example, the Payne-Whitham (PW) model [13], in which the two dependent variables are density and average velocity of the vehicles. Prigogine and Herman [14] developed traffic flow equations based on the Boltzmann equation. These equations were further refined by Paveri-Fontana [15]. Based on Paveri-Fontana's equations, a second order hyperbolic traffic model was defined in [16]. A distinguishing feature of the model in [16] is that it aptly captures the anisotropic behavior of traffic, that is, drivers largely react to the behavior of cars driving ahead of them, as opposed to those driving behind them. Gas dynamic-based two species traffic models with the two species being cars and trucks have also been defined [16], [17]. There have also been papers on analysis of stability in traffic flows [22], [23], and control methods for PDE traffic models [24].

In this paper, we use the model in [16], and within the framework of this model, assume two species of vehicles - namely, malicious and non-malicious (normal) vehicles. The malicious vehicles may be arbitrarily interspersed among the non-malicious vehicles, as schematically shown in Fig 1. Within this PDE framework, a controller is designed, by adopting which the malicious vehicles can create arbitrary velocity/density wave profiles on the highway, and these waves percolate through the non-malicious vehicles. We show how the controller can be extended to represent cyber-attacks on a larger class of second order traffic PDE models, and also in scenarios where the attacker has imperfect information of the traffic parameters. A preliminary version of this paper [25] focused purely on perfect information scenarios.

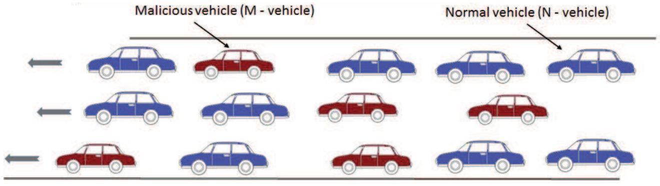

Fig. 1. Malicious $M-$vehicles arbitrarily dispersed among normal $N-$vehicles N on a highway

[1] Meysam Ghanavati is a graduate student in the Department of Electrical Engineering & Computer Science, Wichita State University, Wichita, KS mxghanavati@wichita.edu

[2] Animesh Chakravarthy is a faculty in the Department of Aerospace Engineering and the Department of Electrical Engineering & Computer Science, Wichita State University, Wichita, KS. animesh.chakravarthy@wichita.edu

[3] Prathyush P. Menon is a faculty in the Department of Mathematics, University of Exeter, UK. p.m.prathyush@exeter.ac.uk

This paper is organized as follows. Section II discusses the single-species PDE traffic model [16], and then demonstrates how this model can be extended to analyze a cyber-attack scenario. Section III describes the development of analytical control expressions for two cases: one in which the malicious vehicles generate velocity inputs to modify the spatial velocity profile on the highway, and another in which their generated velocity inputs modify the spatial density profile on the highway. Section IV shows how these malicious waves can be generated even when the attacker has imperfect information of the traffic parameters. Section V contains a discussion on the nature of the velocity/density profiles that the malicious vehicles seek to create. Section VI presents simulation results, while Section VII presents the conclusions.

II. PDE Traffic Model

A. Single Species Model

In this paper, we use a second order macroscopic model proposed in [16]. The development of the model was inspired by earlier gas-kinetic-based models [14], [15]. The macroscopic model has two independent variables x and t, where $x \in [0, L]$ represents a spatial variable and $t \in [0, \infty)$ represents time. The equations for the macroscopic model are:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho V)}{\partial x} = 0 \tag{1}$$

$$\frac{\partial (\rho V)}{\partial t} + \frac{\partial (\rho V^2 + \rho\theta)}{\partial x} = \rho \frac{V^{eq} - V}{\tau} \tag{2}$$

where, $\rho(t, x) : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ represents the average density of vehicles (in vehicles/km/lane), and $\rho(t, x) \in [0, \rho_{max}]$ where ρ_{max} is the maximum possible density. $V(t, x) : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is the average velocity of the vehicles, $\theta(t, x) : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is the velocity variance in the region $[x - dx/2,\ x + dx/2]$. The velocity variance is a function of density and average velocity, $\theta = A(\rho)V^2$, where $A(\rho) = A_0 + \Delta A[\tanh(\frac{\rho-\rho_c}{\Delta\rho}) + 1]$, and A_0, ΔA, ρ_c are positive scalars [16]. $V^{eq}(t, x) : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ represents the average equilibrium velocity and is given by:

$$V^{eq}(t, x) = V^0 - P(\rho_a)B(\delta_v)\rho_a\tau\theta \tag{3}$$

where V^0 is the average desired velocity, τ is the average relaxation time, and ρ_a is the average density computed at the "interaction point". The interaction point x_a is a reference point, as shown in Fig 2, ahead of the "average car" located at x at an instant of time. It is assumed that the average car at x speeds up or slows down depending on the behavior of the average car at x_a. The interaction point is typically defined by $x_a = x + \gamma(l + VT)$, where $l = 1/\rho_{max}$ is the average vehicle length, T is the average time headway and $\gamma \in [1, 3]$ represents an anticipation factor [16]. In (3), the factor $P(\rho_a)$ accounts for the probability of overtaking, as well as the existence of a finite interaction-free space, and is defined as:

$$P(\rho_a) = \frac{V^0 \rho_a T^2}{2\tau A(\rho_{max})(1 - (\rho_a/\rho_{max}))^2} \tag{4}$$

As evident from (4), as ρ_a approches ρ_{max} (that is, the number of vehicles increases), $P(\rho_a)$ becomes progressively larger and this in turn makes $V^{eq}(t, x)$ in (3) smaller, that is, the equilibrium velocity decreases with increasing density. The $B(\delta_v)$ term, defined as [16]:

$$B(\delta_v) = \delta_v \frac{e^{-\delta_v^2/2}}{\sqrt{2\pi}} + (1 + \delta_v^2) \int_{-\infty}^{\delta_v} dy \frac{e^{-y^2/2}}{\sqrt{2\pi}}, \tag{5}$$

takes into account the anisotropic effects of the traffic flow, that is, the driver of a car responds more to the traffic ahead, than the traffic behind. In (5), the term $\delta_v = (V - V_a)/\sqrt{\theta + \theta_a}$ is a dimensionless velocity difference between the average car at x and that at its interaction point x_a. Terms V_a and θ_a denote the average velocity and velocity variance computed at the interaction point x_a. In (3), V^{eq} thus represents a balance between the desire of the average car at x to drive at it's preferred speed V^0 and the slowing down effect it experiences due to its interaction with the average car at x_a.

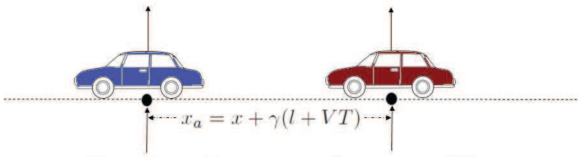

Fig. 2. Concept of *interaction point* and *average car*.

While we primarily utilize the above model [16], the methods developed in this paper are applicable to a broader class of second order traffic PDE models, comprising (1) and (6). While (1) is a standard equation used for any traffic model, (2) represents a special case of a large class of PDE models (including for example, [13], [18], [19], [20], [21] among many others) with the following general structure:

$$\frac{\partial (\rho V)}{\partial t} + \frac{\partial (\rho V^2)}{\partial x} + \frac{\partial \mathcal{P}}{\partial x} = \rho \frac{V^{eq} - V}{\tau} \tag{6}$$

where $\mathcal{P}$ represents the traffic pressure. For the model proposed in [16], the traffic pressure is inferred based on empirical data as $\mathcal{P} = \rho\theta = \rho A(\rho)V^2$, as given in (2).

B. Two-Species Model

The model in (1)-(2) is extended to a two-species model with subscripts M and N, as schematically shown in Fig 1. The dynamics of the proposed two-species model are:

$$\frac{\partial \rho_N}{\partial t} + \frac{\partial \left(\rho_N V_N\right)}{\partial x} = 0 \tag{7}$$

$$\frac{\partial \left(\rho_N V_N\right)}{\partial t} + \frac{\partial \left(\rho_N V_N^2 + \rho_N \theta_N\right)}{\partial x} = \rho_N \frac{V_N^{eq} - V_N}{\tau} \tag{8}$$

$$\frac{\partial \rho_M}{\partial t} + \frac{\partial \left(\rho_M V_M\right)}{\partial x} = 0 \tag{9}$$

$$\frac{\partial \left(\rho_M V_M\right)}{\partial t} + \frac{\partial \left(\rho_M V_M^2 + \rho_M \theta_M\right)}{\partial x} = \rho_M \frac{V_M^{eq} - V_M}{\tau} \tag{10}$$

In (7)-(10), the average equilibrium velocities of the malicious and normal vehicles, V_M^{eq} and V_N^{eq} are defined as follows:

$$V_M^{eq} = V_M^0 - P(\rho_{aM}, \rho_{aN})\tau(B_{MN}\rho_N\theta_N + B_{MM}\rho_M\theta_M) \tag{11}$$

$$V_N^{eq} = V_N^0 - P(\rho_{aM}, \rho_{aN})\tau(B_{NM}\rho_M\theta_M + B_{NN}\rho_N\theta_N) \tag{12}$$

In (11)-(12), in general $B_{ij} := B(\delta_{v,ij})$, where

$$\delta_{v,ij} = (V_i - V_{a,j})/\sqrt{\theta_i + \theta_{a,j}}\ , i = M, N;\ j = M, N$$

represents the dimensionless velocity difference between the vehicle at location x and the vehicle at interaction point x_a as shown in figure 2. In (11), the equilibrium velocity V_M^{eq} of the malicious vehicles is given by the balance between the desire of the "average" malicious vehicle at x to drive at it's preferred velocity V_M^0, and the slowing down effects it experiences due to the average malicious vehicle at x_a, and the average normal vehicle at x_a. A corresponding statement can be made for the equilibrium velocity V_N^{eq} of the normal vehicles. The effects of interaction between the two species are thus manifested in the V_M^{eq} and V_N^{eq} terms. The PDEs (8) and (10) essentially govern the spatio-temporal evolution of the velocities V_N and V_M, to these equilibrium velocities V_M^{eq} and V_N^{eq}, respectively. The variable $P(\rho_M, \rho_N)$ is defined as follows:

$$P(\rho_M, \rho_N) = \frac{T^2\left(\rho_M V_M^0 + \rho_N V_N^0\right)}{2\tau A(\rho_{max})\left(1 - \frac{\rho_M + \rho_N}{\rho_{max}}\right)^2} \tag{13}$$

Therefore, the average equilibrium velocity of normal vehicles in (12) can be written as follows:

$$\begin{aligned} V_N^{eq} &= V_N^0\left(1 - \frac{T^2\rho_{aN}\left(B_{NM}\rho_M\theta_M + B_{NN}\rho_N\theta_N\right)}{2A(\rho_{max})\left(1 - \frac{\rho_{aM}+\rho_{aN}}{\rho_{max}}\right)^2}\right) \\ &- V_M^0\left(\frac{T^2\rho_{aM}\left(B_{NM}\rho_M\theta_M + B_{NN}\rho_N\theta_N\right)}{2A(\rho_{max})\left(1 - \frac{\rho_{aM}+\rho_{aN}}{\rho_{max}}\right)^2}\right) \end{aligned} \tag{14}$$

We see that V_N^{eq} is affine in V_M^0, and can be rewritten as:

$$V_N^{eq} = \alpha_N V_M^0 + \beta_N \tag{15}$$

The term V_M^0 thus influences the average equilibrium velocity of the normal vehicles, and may be viewed as a control input for the two-species model in (7)-(10).

III. Reference Tracking Control

The objective is to vary (in space and time) the average velocity of the malicious vehicles in (9)-(10), which then can induce undesirable waves in the spatial velocity and density profiles of the normal vehicles in (7)-(8). In other words, by performing these velocity modulations, the malicious vehicles force changes in the velocity and density profiles of the normal vehicles, till they are modified to track certain pre-specified, reference profiles. A discussion on examples of the kind of reference profiles, $V_N^r(t,x)$ and $\rho_N^r(t,x)$, the malicious vehicles may choose, is provided in Section V.

A. Velocity Profile Manipulation

A Lyapunov-based control law is designed, using which, the malicious vehicles can generate a malicious velocity profile on the highway. Towards this end, consider a Lyapunov functional Z_V as follows:

$$Z_V(t) \equiv \frac{1}{2}\int_0^L \left(V_N(t,x) - V_N^r(t,x)\right)^2 \mathrm{d}x \tag{16}$$

where V_N^r represents the smooth bounded reference velocity profile that the malicious vehicles seek to create among the normal vehicles, and L represents the length of the highway under consideration. Based on the theory of Lyapunov functions [26] and considering V_M^0 to be the control input for the malicious vehicles, if V_M^0 is chosen such that $\frac{\partial Z_V}{\partial t}$ is negative definite, then it is guaranteed that the velocity of the normal vehicles will asymptotically track V_N^r. Substituting $\frac{\partial \rho_N}{\partial t}$ from (7) in (8), and using (14), we arrive at:

$$\begin{aligned} \frac{\partial V_N}{\partial t} &= -\frac{V_N}{\rho_N}\frac{\partial\left(\rho_N V_N\right)}{\partial x} - \frac{1}{\rho_N}\frac{\partial\left(\rho_N V_N^2 + \rho_N\theta_N\right)}{\partial x} \\ &+ \frac{V_N^{eq} - V_N}{\tau} \equiv F_{V_N} + \frac{\alpha_N}{\tau}V_M^0 \end{aligned} \tag{17}$$

where, α_N can be inferred from (14)-(15), and F_{V_N} is:

$$\begin{aligned} F_{V_N} &= -\frac{V_N}{\rho_N}\frac{\partial\left(\rho_N V_N\right)}{\partial x} - \frac{1}{\rho_N}\frac{\partial\left(\rho_N V_N^2 + \rho_N\theta_N\right)}{\partial x} \\ &- \frac{V_N}{\tau} + \frac{\beta_N}{\tau} \end{aligned} \tag{18}$$

To attain the objective that V_N tracks V_N^r, we take the time derivative of the Lyapunov function (16), substitute (17) in this time derivative, and finally obtain the following equation:

$$\frac{\partial Z_V}{\partial t} = \int_0^L e_v\left(F_{V_N} + \frac{\alpha_N}{\tau}V_M^0 - \frac{\partial V_N^r(t,x)}{\partial t}\right)\mathrm{d}x \tag{19}$$

where, $e_v(t,x) \equiv (V_N(t,x) - V_N^r(t,x))$. If we define V_M^0 as

$$V_M^0 = \frac{\tau}{\alpha_N}\left(-F_{V_N} + \frac{\partial V_N^r(t,x)}{\partial t} - K_v e_v(t,x)\right) \tag{20}$$

then, the time derivative of the Lyapunov function is:

$$\frac{\partial Z_V}{\partial t} = -\int_0^L K_v e_v(t,x)^2 \mathrm{d}x \tag{21}$$

With $K_v > 0$ in (21), we have that $\frac{\partial Z_V}{\partial t} < 0$, which in turn means that the average velocity of the normal vehicles will globally exponentially track the reference velocity profile. Since V_M^0 in (20) is a function of α_N, and (14)-(15) shows that α_N depends on ρ_M, therefore as long as $\rho_M \neq 0\ \forall x$, α_N remains non-zero, thereby ensuring that V_M^0 remains bounded.

The smaller the magnitude of K_v in (21), the more gradual the decay of Z_V, and consequently, the more subtle the velocity changes required by the malicious vehicles to meet their objective. By proper choice of K_v and time-varying reference $V_N^r(t,x)$, it can be ensured that the generated velocity profile for the malicious vehicles is always physically feasible.

B. Density Profile Manipulation by the Malicious Vehicles

In this section a control law is designed, using which the malicious vehicles can cause the density profile of the normal vehicles to track a pre-defined reference $\rho_N^r(t,x)$. Define a Lyapunov function Z_ρ as follows:

$$Z_\rho(t) \equiv \frac{1}{2}\int_0^L \left(\rho_N(t,x) - \rho_N^r(t,x)\right)^2 \mathrm{d}x \tag{22}$$

We seek to determine $V_M^0(t,x)$ such that the time derivative of Z_ρ is negative definite, as this will guarantee that the density of the normal vehicles $\rho_N(t,x)$ tracks $\rho_N^r(t,x)$. The time derivative of (22) along the system trajectories (7)-(10) is:

$$\frac{\partial Z_\rho}{\partial t} = \int_0^L e_\rho(t,x)\left(\frac{\partial \rho_N(t,x)}{\partial t} - \frac{\partial \rho_N^r(t,x)}{\partial t}\right)\mathrm{d}x \tag{23}$$

where, $e_\rho(t,x) \equiv (\rho_N(t,x) - \rho_N^r(t,x))$. Now if we define $V_M^0(t,x)$ in a way such that $\frac{\partial \rho_N(t,x)}{\partial t}$ is as follows:

$$\frac{\partial \rho_N(t,x)}{\partial t} = \frac{\partial \rho_N^r(t,x)}{\partial t} - K_\rho e_\rho(t,x) \tag{24}$$

then, with a choice of $K_\rho > 0$, the Lyapunov function (22) becomes a decreasing function of time, as desired. Substituting (24) in (7), and integrating with respect to x, we arrive at:

$$\rho_N(t,x)V_N(t,x) = \int_0^x \left(-\frac{\partial \rho_N^r(t,x)}{\partial t} + K_\rho e_\rho(t,x) \right) \mathrm{d}x + C \tag{25}$$

where C is a constant of integration. Taking $C = \rho_N(t,0)V_N(t,0)$, and substituting the above equation in (8):

$$\rho_N \frac{V_N^{eq} - V_N}{\tau} = \frac{\partial Q}{\partial t} + \frac{\partial \left(\rho_N V_N^2 + \rho_N \theta_N\right)}{\partial x} \tag{26}$$

where, $Q(t,x)$ in the above equation is given by:

$$Q = \int_0^x \left(-\frac{\partial \rho_N^r(t,x)}{\partial t} + K_\rho e_\rho \right) \mathrm{d}x + \rho_N(t,0)V_N(t,0) \tag{27}$$

Substituting V_N^{eq} from (14)-(15) into (26), the requisite V_M^0 is:

$$\begin{aligned} V_M^0(t,x) &= \left(\frac{\tau}{\rho_N \alpha_N}\right) \Big[\frac{\partial \left(\rho_N V_N^2 + \rho_N \theta_N\right)}{\partial x} \\ &- \frac{\rho_N \beta_N}{\tau} + \frac{\rho_N V_N}{\tau} + \frac{\partial Q}{\partial t} \Big] \end{aligned} \tag{28}$$

This $V_M^0(t,x)$ ensures that $\frac{\partial Z_\rho}{\partial t}$ is negative definite, which in turn, ensures that $\rho_N(t,x)$ asymptotically tracks $\rho_N^r(t,x)$.

C. *Velocity/Density Manipulation for general PDE model*

For the general form of the traffic PDE model defined by (1) and (6), a controller can be derived along similar lines. For density profile manipulation, defining Lyapunov function to be the same as in (22), and considering different $\mathcal{P}_N$ for different models, while assuming V_N^{eq} can be written as an affine function of V_M^0 (that is, V_M^0 has the structure of (15) but with a possibly different α_N and β_N), the desired density profile can be achieved by choosing V_M^0 as

$$\begin{aligned} V_M^0(t,x) &= \left(\frac{\tau}{\rho_N \alpha_N}\right) \Big[\frac{\partial \left(\rho_N V_N^2 + \mathcal{P}_N\right)}{\partial x} \\ &- \frac{\rho_N \beta_N}{\tau} + \frac{\rho_N V_N}{\tau} + \frac{\partial Q}{\partial t} \Big] \end{aligned} \tag{29}$$

where Q is defined as (27). Along similar lines, the expression for V_M^0 that will achieve the desired velocity profile in the general traffic model can be obtained.

IV. Imperfect Information Scenarios

In this section, we consider imperfect information scenarios, wherein the attacker does not have access to the values of traffic parameters, such as the relaxation time constant τ, and average time headway T. These parameters are considered to be uncertain and are assumed to be $\hat{\tau}$ and $\hat{T}$ for analysis. We analyze the possibilities of what can happen if the attacker uses values $\hat{\tau}$ and $\hat{T}$, that are different from their respective true values τ and T, in both - the velocity profile manipulation, as well as density profile manipulation cases.

A. *Velocity Profile Manipulation with Imperfect Information*

When the attacker employs assumed values $\hat{\tau}$ and $\hat{T}$ in the V_M^0 of (20), then the time derivative of (16) along the trajectories of (7)-(10), assumes the form:

$$\dot{Z}_V(t) = -\frac{\hat{\tau} T^2}{\tau \hat{T}^2} K_v Z_V(t) + g_V \tag{30}$$

where the term g_V is as follows:

$$\begin{aligned} g_V = &\left(1 - \frac{\hat{\tau} T^2}{\tau \hat{T}^2}\right) \int_0^L \left(\bar{F}_{V_N}(t,x) - \frac{\partial V_N^r(t,x)}{\partial t} \right) e_v(t,x) \mathrm{d}x \\ &- \left(1 - \frac{T^2}{\hat{T}^2}\right) \int_0^L \left(\frac{-V_N(t,x) + V_N^0}{\tau} \right) e_v(t,x) \mathrm{d}x \end{aligned} \tag{31}$$

Since τ and T are positive scalars, it can be seen from (30) that as long as $\hat{\tau}$ and $\hat{T}$ are positive, Z_V remains stable when $g_V = 0$. However, a non-zero g_V may influence the asymptotic stability of Z_V, that is, it may prevent Z_V from asymptotically decaying to zero. This can happen under conditions that are determined as follows. It can be observed from (16) and (31) that when $e_v(t,x) = 0, \forall x \in [0,L]$, (that is, $V_N(t,x) = V_N^r(t,x), \forall x \in [0,L]$), then both Z_V and g_V are zero. In (30), g_V thus has the structure of a vanishing perturbation, that is, $Z_V = 0 \Rightarrow g_V = 0$.

Define another Lyapunov function $\mathcal{V} = \frac{1}{2} Z_V^2$. Then, $\mathcal{V}$ satisfies the following conditions:

$$\begin{aligned} c_1 |Z_V|^2 \leq \mathcal{V}(Z_V) &\leq c_2 |Z_V|^2 \\ \frac{\partial \mathcal{V}}{\partial Z_V} \left(-\frac{\hat{\tau} T^2}{\tau \hat{T}^2} K_v Z_V \right) &\leq -c_3 |Z_V|^2 \\ \left| \frac{\partial \mathcal{V}}{\partial Z_V} \right| &\leq c_4 |Z_V| \end{aligned} \tag{32}$$

In (32), $c_1 = c_2 = \frac{1}{2}$, $c_3 = \frac{T^2 \hat{\tau}}{\hat{T}^2 \tau} K_v$, $c_4 = 1$. Since g_V is a vanishing perturbation, $Z_V(t)$ is globally exponentially stable if $|g_V| \leq \frac{c_3}{c_4} |Z_V|$ [26]. Thus, as long as

$$|g_V| \leq K_v \frac{T^2 \hat{\tau}}{\hat{T}^2 \tau} |Z_V| \tag{33}$$

is satisfied, the controller will guarantee that the reference velocity profile is tracked. Note that the upper bound in (33) is conservative. It is apparent from (33) that by increasing the value of K_v, the upper bound on $|g_V|$ with which exponential stability of (30) is guaranteed, can be increased. However, too large a K_v can require larger velocity changes to be performed by the malicious vehicles, which the attacker may not always desire, since it would make the attack less stealthy.

When perfect information of τ and T is not available, the same can be determined by adding an adaptation law to the existing control laws, thereby creating an adaptive cyber-attack system. Towards this end, we ignore the effect of T on the B_{NN} and B_{NM} terms. This is justified because for those portions of the highway where the traffic is smooth, $V \approx V_a$, which makes B_{NN} and B_{NM} independent of T. Define new variables $h = T^2$, $w = \tau / h$, and error terms $\tilde{h}$ and $\tilde{w}$, as $\tilde{h} \equiv h - \hat{h}$, and $\tilde{w} \equiv w - \hat{w}$, where $\hat{h}$ and $\hat{w}$ represent estimates of h and w, respectively. Then, V_N^{eq} in (14) can be written as:

$$V_N^{eq} = V_N^0 + h\beta_{TN} + h\alpha_{TN} V_M^0 \tag{34}$$

where, β_{TN} and α_{TN} are both independent of τ and T. Using constants $\gamma_1 > 0$, $\gamma_2 > 0$, we define a Lyapunov function as:

$$\bar{Z}_V(t) \equiv \frac{1}{2}\int_0^L \left(V_N(t,x) - V_N^r(t,x)\right)^2 \mathrm{d}x + \frac{\gamma_1}{2\tau}\tilde{h}^2 + \frac{\gamma_2}{2w}\tilde{w}^2 \tag{35}$$

The associated control law is updated to:

$$\begin{aligned} V_M^0 &= \frac{\hat{\tau}}{\hat{h}\alpha_{TN}}\Bigg(- \bar{F}_{V_N} - \frac{-V_N + V_N^0}{\hat{\tau}} - \frac{\hat{h}\beta_{TN}}{\hat{\tau}} \\ &+ \frac{\partial V_N^r(t,x)}{\partial t} - K_v e_v\Bigg) \end{aligned} \tag{36}$$

where $\bar{F}_{V_N} = -\frac{V_N}{\rho_N}\frac{\partial(\rho_N V_N)}{\partial x} - \frac{1}{\rho_N}\frac{\partial\left(\rho_N V_N^2 + \rho_N\theta_N\right)}{\partial x}$

Then, by taking the time derivative of (35) and substituting V_M^0 from (36), we arrive at:

$$\begin{aligned} \frac{\partial \bar{Z}_V}{\partial t} = &-K_v\frac{\hat{w}}{w}\int_0^L e_v(t,x)^2\mathrm{d}x + \left(\frac{\tilde{w}}{w}\right) \\ &\times\left[\int_0^L \left(\bar{F}_{V_N}(t,x) - \frac{\partial V_N^r(t,x)}{\partial t}\right) e_v\mathrm{d}x - \gamma_2\dot{\hat{w}}\right] \\ &- \left(\frac{\tilde{h}}{\tau}\right)\left[\int_0^L \left(\frac{-V_N(t,x) + V_N^0}{\hat{h}}\right) e_v\mathrm{d}x + \gamma_1\dot{\hat{h}}\right] \end{aligned} \tag{37}$$

It can then be guaranteed that $\frac{\partial \bar{Z}_V}{\partial t} < 0$ if the terms in square brackets that multiply $\frac{\tilde{w}}{w}$ and $\frac{\tilde{h}}{\tau}$ are each made equal to zero. This can be achieved by defining $\dot{\hat{h}}$ and $\dot{\hat{w}}$ as follows:

$$\dot{\hat{h}} = \frac{-1}{\gamma_1}\int_0^L \left(\frac{-V_N(t,x) + V_N^0}{\hat{h}}\right) e_v(t,x)\mathrm{d}x \tag{38}$$

$$\dot{\hat{w}} = \frac{1}{\gamma_2}\int_0^L \left(\bar{F}_{V_N}(t,x) - \frac{\partial V_N^r(t,x)}{\partial t}\right) e_v(t,x)\mathrm{d}x \tag{39}$$

Eqns (36), (38) and (39) thus represent adaptive laws which an attacker may use in imperfect information scenarios, and successfully achieve manipulation of the velocity profile.

B. Density Profile Manipulation with Imperfect Information

We now look at the effects of imperfect values $\hat{\tau}$ and $\hat{T}$ on an attempt of the attacker to manipulate the highway density profile. Using these imperfect values in the V_M^0 of (28), the time derivative of (22) along the trajectories of (7)-(10) is:

$$\dot{Z}_\rho(t) = -\frac{\hat{\tau}T^2}{\tau\hat{T}^2}K_\rho Z_\rho(t) + g_\rho \tag{40}$$

where, g_ρ is as follows:

$$g_\rho = \frac{\tilde{w}}{w}\tilde{f}_w + \frac{\tilde{h}}{h}\tilde{f}_h \tag{41}$$

and the expressions for $\tilde{f}_w$ and $\tilde{f}_h$ are lengthy (and therefore omitted), but straightforward to derive. g_ρ has the structure of a vanishing perturbation, that is, $Z_\rho = 0 \Rightarrow g_\rho = 0$. Doing an analysis similar to that performed in the preceding subsection, it can be shown that the attacker can still achieve exponential tracking of the density profile as long as g_ρ satisfies:

$$|g_\rho| \leq K_\rho\frac{\hat{w}}{w}|Z_\rho| \tag{42}$$

As before, the attacker can choose a larger value of K_ρ to ensure that g_ρ satisfies (42), but the larger K_ρ may come at the expense of the malicious vehicles having to perform larger velocity changes, which the attacker may not desire. One may then define a modified Lyapunov function $\bar{Z}_\rho(t)$ similar to $\bar{Z}_V(t)$ in (35), but with $V_N(t,x)$ and $V_N^r(t,x)$ replaced by $\rho_N(t,x)$ and $\rho_N^r(t,x)$, respectively. Correspondingly, the control law is modified as follows:

$$\begin{aligned} V_M^0(t,x) &= \left(\frac{\hat{\tau}}{\hat{h}\rho_N\alpha_{TN}}\right)\Big[\frac{\partial Q}{\partial t} + \frac{\partial\left(\rho_N V_N^2 + \rho_N\theta_N\right)}{\partial x} \\ &- \frac{\rho_N V_N^0 + \hat{h}\rho_N\beta_{TN}}{\hat{\tau}} + \frac{\rho_N V_N}{\hat{\tau}}\Big] \end{aligned} \tag{43}$$

where Q is as given in (27). Taking the time derivative of $\bar{Z}_\rho(t)$, and using (7), we arrive at:

$$\frac{\partial \bar{Z}_\rho}{\partial t} = -K_\rho\frac{\hat{w}}{w}\int_0^L e_\rho^2\mathrm{d}x + \frac{\tilde{w}}{w}(\tilde{f}_w - \gamma_2\dot{\hat{w}}) - \frac{\tilde{h}}{\tau}(\frac{\tilde{f}_h}{\hat{h}} + \gamma_1\dot{\hat{h}}) \tag{44}$$

Tracking error stability is guaranteed if $\frac{\partial \bar{Z}_\rho}{\partial t} < 0$ is achieved, and the same can be obtained by defining $\dot{\hat{h}}$ and $\dot{\hat{w}}$ as follows:

$$\dot{\hat{h}} = \frac{-\tilde{f}_h}{\hat{h}\gamma_1}, \quad \dot{\hat{w}} = \frac{\tilde{f}_w}{\gamma_2} \tag{45}$$

Eqns (43) and (45) thus represent adaptive laws which an attacker may use in imperfect information scenarios, and successfully achieve manipulation of the density profile.

C. Velocity/Density Manipulation for general PDE model

In this section, the objective is to manipulate the velocity and density profiles of the normal vehicles, while using the general form of traffic model (1) and (6), assuming the attacker does not have access to the exact value of τ. We note that the general form of the PDE model does not include the T term. Assuming V_N^{eq} has the form of (15), that is, V_N^{eq} is affine in V_M^0 and with a possibly different α_N and β_N from that given in (14), we proceed as follows. Define the error between the estimate and actual values of τ by $\tilde{\tau} \equiv \tau - \hat{\tau}$, where $\hat{\tau}$ represents the estimate of τ. For velocity profile manipulation, the Lyapunov function and the related control law are:

$$\bar{Z}_V(t) \equiv \frac{1}{2}\int_0^L \left(V_N(t,x) - V_N^r(t,x)\right)^2 \mathrm{d}x + \frac{\gamma}{2\tau}\tilde{\tau}^2 \tag{46}$$

$$V_M^0 = \frac{\hat{\tau}}{\alpha_N}\left(- \bar{F}_{V_N} + \frac{V_N - \beta_N}{\hat{\tau}} - K_v e_v(t,x)\right) \tag{47}$$

where, $\bar{F}_{V_N} = -\frac{V_N}{\rho_N}\frac{\partial(\rho_N V_N)}{\partial x} - \frac{1}{\rho_N}\frac{\partial\left(\rho_N V_N^2 + \mathcal{P}_N\right)}{\partial x} - \frac{\partial V_N^r(t,x)}{\partial t}$, and $\gamma > 0$ is a constant. Taking the derivative of (46) with respect to t and substituting V_M^0 from (47), we arrive at:

$$\frac{\partial \bar{Z}_V}{\partial t} = -K\frac{\hat{\tau}}{\tau}\int_0^L e_v^2\mathrm{d}x + \left(\frac{\tilde{\tau}}{\tau}\right)\left[\int_0^L \bar{F}_{V_N} e_v\mathrm{d}x - \gamma\dot{\hat{\tau}}\right] \tag{48}$$

Stability of the controller is guaranteed by defining $\dot{\hat{\tau}}$ as:

$$\dot{\hat{\tau}} = \frac{1}{\gamma}\int_0^L \bar{F}_{V_N}(t,x)e_v(t,x)\mathrm{d}x \tag{49}$$

Eqns (47) and (49) thus represent an adaptive cyber-attack system for velocity profile manipulation. For density manipulation, the modified Lyapunov function and control law are:

$$\bar{Z}_\rho(t) \equiv \frac{1}{2}\int_0^L \left(\rho_N(t,x) - \rho_N^r(t,x)\right)^2 \mathrm{d}x + \frac{\gamma}{2\tau}\tilde{\tau}^2 \quad (50)$$

$$V_M^0(t,x) = \left(\frac{\hat{\tau}}{\rho_N \alpha_N}\right)\Big[\frac{\partial\left(\rho_N V_N^2 + \mathcal{P}_N\right)}{\partial x} - \frac{\rho_N \beta_N}{\hat{\tau}} + \frac{\rho_N V_N}{\hat{\tau}} + \frac{\partial Q}{\partial t}\Big] \quad (51)$$

Taking the derivative of (50) with respect to t, and using (7), (43), after some lengthy calculations, we eventually arrive at:

$$\frac{\partial \bar{Z}_\rho}{\partial t} = -K_\rho \frac{\hat{\tau}}{\tau}\int_0^L e_\rho(t,x)^2 \mathrm{d}x - \frac{\gamma}{\tau}\tilde{\tau}\dot{\hat{\tau}} + \int_0^L e_\rho(t,x) \left(\frac{\partial}{\partial x}\int_0^t \Big[\frac{\partial\left(\rho_N V_N^2 + \mathcal{P}_N\right)}{\partial x} + \frac{\partial \rho_N^r}{\partial t}\Big]\left(1 - \frac{\hat{\tau}}{\tau}\right)\mathrm{d}t\right)\mathrm{d}x \quad (52)$$

$$\frac{\partial \bar{Z}_\rho}{\partial t} = -K_\rho \frac{\hat{\tau}}{\tau}\int_0^L e_\rho(t,x)^2 \mathrm{d}x - \frac{\gamma}{\tau}\tilde{\tau}\dot{\hat{\tau}} + \left(1 - \frac{\hat{\tau}}{\tau}\right)\int_0^L e_\rho(t,x) \left(\frac{\partial}{\partial x}\int_0^t \frac{\partial\left(\rho_N V_N^2 + \mathcal{P}_N\right)}{\partial x} + \frac{\partial \rho_N^r}{\partial t}\mathrm{d}t\right)\mathrm{d}x \quad (53)$$

It can be guaranteed that the density profile of the normal vehicles will track the reference density profile, that is, $\frac{\partial \bar{Z}_\rho}{\partial t} < 0$ if $\hat{\tau}$ is determined from the following adaptation law:

$$\dot{\hat{\tau}} = \frac{1}{\gamma}\int_0^L e_\rho(t,x) \quad \frac{\partial}{\partial x}\int_0^t \frac{\partial\left(\rho_N V_N^2 + \mathcal{P}_N\right)}{\partial x} + \frac{\partial \rho_N^r}{\partial t}\mathrm{d}t\Bigg)\mathrm{d}x \quad (54)$$

Eqns (51) and (54) thus represent an adaptive cyber-attack system for density profile manipulation.

V. Choice of reference velocity and density profiles

As mentioned earlier, the reference density and velocity profiles chosen by the malicious vehicles are such that once they have formed, they spontaneously evolve to form shock waves on the highway. It is assumed that once ρ_N^r (V_N^r) is attained, the malicious vehicles exit the highway stretch under consideration, and this exit occurs with a time constant that is small enough so that the exit can be considered instantaneous when compared to the time scale of the macroscopic model. Thus once the ρ_N^r (V_N^r) reference profile is attained, the two-species model reverts to a single-species model, which now contains only normal vehicles. We can therefore employ a single-species analysis to determine the ρ_N^r (V_N^r) that the malicious vehicles choose, with the intent of generating a shock wave in the subsequent (single species) traffic. This analysis involves the use of characteristic velocities [18].

We consider the reference density profile qualitatively depicted in Fig 3, and quantified by the parameters ρ_{min}, ρ_{max}, $x_{\rho min}$, $x_{\rho max}$. By appropriate selection of values of these parameters, a unique shock of the desired magnitude, speed and location (at which it first forms) can be obtained. This is done using the method of characteristics. Characteristic curves are specific curves on the $x - t$ plane, along which the PDEs are transformed into ordinary differential equations (ODEs). The single species PDE model is rewritten as follows:

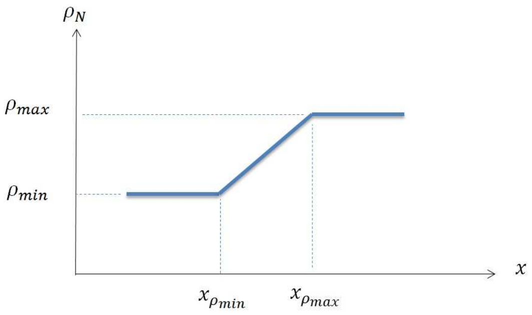

Fig. 3. Ansatz for reference density profile

$$\frac{\partial U}{\partial t} + B(U)\frac{\partial U}{\partial x} = H(U) \quad (55)$$

where, $U = [u_1, u_2]'$, with $u_1 \equiv \rho_N$, $u_2 \equiv \rho_N V_N$, and:

$$B(U) = \begin{bmatrix} 0 & 1 \\ -(\frac{u_2}{u_1})^2(1 + A(u_1)) + \frac{u_2^2}{u_1}A'(u_1) & 2\frac{u_2}{u_1}(1 + A(u_1)) \end{bmatrix} \quad (56)$$

where $'$ denotes derivative with respect to u_1. The eigenvalues of $B(U)$ define the slopes of the characteristic curves and are:

$$\lambda_i(B) = \frac{u_2(A \pm \sqrt{A^2 + A + A'u_1} + 1)}{u_1}, \quad i = 1,2 \quad (57)$$

The system of two equations represented in (55) can be combined into a single second order equation as follows.

$$\begin{aligned} -\frac{\partial^2 u_1}{\partial t^2} \quad &+ \quad \left(\frac{u_2^2}{u_1^2}(A'u_1 - A - 1)\right)\frac{\partial^2 u_1}{\partial x^2} - 2u_2 f(u_1)\frac{\partial^2 u_1}{\partial t \partial x} \\ &- \quad \frac{4u_2}{u_1^2}(A'u_1 - A - 1)\frac{\partial^2 u_1}{\partial x \partial t} + 2f(u_1)\left(\frac{\partial^2 u_1}{\partial t \partial x}\right)^2 \\ &+ \quad \left(\frac{u_2^2}{u_1^3}(A''u_1^2 - 2A'u_1 + 2A + 2)\right)\left(\frac{\partial u_1}{\partial x}\right)^2 \\ &= \quad \frac{1}{\tau}\frac{\partial(u_1 V^{eq})}{\partial x} + \frac{1}{\tau}\frac{\partial u_1}{\partial t} \end{aligned} \quad (58)$$

where, $f(u_1) = (1 + A(u_1))/u_1$. To determine the relative effects of the different parameters, we write (58) in a non-dimensional form. Towards this end, we define $\tilde{t} = t/\bar{T}$, $\tilde{x} = x/L$, $\tilde{V}^{eq} = V^{eq}/V^o$and $\tilde{u}_1 = u_1/u_m$, where u_m is the maximum value of u_1. Here, $\bar{T}$ represents the characteristic time. By choosing $\bar{T} = L/V^o$, (58) assumes the form:

$$\begin{aligned} &- \frac{V^o}{L}\frac{\partial^2 \tilde{u}_1}{\partial \tilde{t}^2} + \left(\frac{u_2^2(A'u_1 - A - 1)}{u_1^2 V^o L}\right)\frac{\partial^2 \tilde{u}_1}{\partial \tilde{x}^2} - \frac{2u_2 f(u_1)}{L}\frac{\partial^2 \tilde{u}_1}{\partial \tilde{t}\partial \tilde{x}} \\ &- \frac{4u_2 u_m(A'u_1 - A - 1)}{u_1^2 L}\frac{\partial^2 \tilde{u}_1}{\partial \tilde{x}\partial \tilde{t}} + \frac{2u_m V^o f(u_1)}{L^3}\left(\frac{\partial^2 \tilde{u}_1}{\partial \tilde{t}\partial \tilde{x}}\right)^2 \\ &+ \left(\frac{u_2^2 u_m}{u_1^3 V^o L}(A''u_1^2 - 2A'u_1 + 2A + 2)\right)\left(\frac{\partial \tilde{u}_1}{\partial \tilde{x}}\right)^2 \\ &= \frac{1}{\tau}\left(\frac{\partial(\tilde{u}_1 \tilde{V}^{eq})}{\partial \tilde{x}} + \frac{\partial \tilde{u}_1}{\partial \tilde{t}}\right) \end{aligned} \quad (59)$$

Multiplying both sides of (59) by τ, the equation assumes the following structure:

$$\tau(LHS) - \left(\frac{\partial(\tilde{u}_1\tilde{V}^{eq})}{\partial\tilde{x}} + \frac{\partial\tilde{u}_1}{\partial\tilde{t}}\right) = 0 \qquad (60)$$

where LHS represents the left hand side of (59). Note that the coefficients of all the derivative terms in (59) are dimensionless, and thus provide a convenient way of determining the relative influence of the different parameters. It is evident that the coefficient of the second order time derivative is $-\frac{V^o}{L}\tau$, and for typical parameter values $V^0 = 110\,\mathrm{kmph}$, $L = 10\,\mathrm{km}$, and $\tau = 15\,\mathrm{sec}$, we obtain $\tau\left|-\frac{V^o}{L}\right| = 0.046$, which is significantly smaller than unity. Using singular perturbation analysis, we can therefore surmise that the dynamics of (60) comprises of two time scales: a fast time scale during which the term LHS decays to zero, followed by a slower time scale during which time the dynamics of (58) can be well-approximated by the equation:

$$\frac{\partial(u_1 V^{eq})}{\partial x} + \frac{\partial u_1}{\partial t} = 0 \qquad (61)$$

The above equation has a characteristic speed given by $\lambda = V^{eq} + u_1\frac{\partial V^{eq}}{\partial u_1}$. Since $u_1 = \rho$, we can see that the characteristic speed is a function of ρ. Thus, for an equilibrium velocity V^{eq} as follows [16]:

$$V^{eq}(\rho) = \frac{\left(-1 + \sqrt{1 + 4V^0\tau\rho A(\rho)P(\rho)}\right)}{2\tau\rho A(\rho)P(\rho)} \qquad (62)$$

we can substitute (62) in (61), and then obtain a typical plot of the characteristic velocity λ as a function of ρ in Fig 4. Fig 4 shows that there is a range of ρ for which $\lambda > 0$, and a range of ρ for which $\lambda < 0$. This trend is similar to that demonstrated by other equilibrium velocity profiles in the literature, and can be used to determine whether an initial condition develops into a shock or otherwise. For example, as seen from Fig 4, for $\rho < 37.5$, we have $\lambda > 0$, while for $\rho > 37.5$, we have $\lambda < 0$. Now consider a scenario wherein at a given time t_1, $\rho_{max}(t_1, x_{\rho_{max}}) > 37.5$, and $\rho_{min}(t_1, x_{\rho_{min}}) < 37.5$. Then, that portion of the density profile with $\rho < 37.5$ will travel forward (since $\lambda > 0$) along the highway, while that portion of the density profile with $\rho > 37.5$ will travel backward (since $\lambda < 0$). This will cause the density profile to become progressively steeper, eventually leading to a shock, which originates between $x_{\rho_{min}}$ and $x_{\rho_{max}}$. The exact location at which the shock first forms is readily determined as follows. At any given time t, the locations of the travelling wavefronts corresponding to ρ_{max} and ρ_{min} are given by:

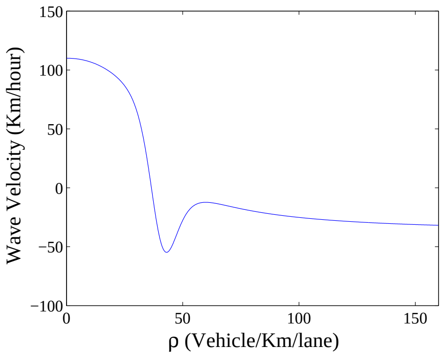

Fig. 4. Characteristic Velocity as a function of density

$$\begin{aligned} x_{\rho_{max}}(t) &= \lambda(\rho_{max}) \times (t - t_1) + x_{\rho_{max}}(t_1) & (63)\\ x_{\rho_{min}}(t) &= \lambda(\rho_{min}) \times (t - t_1) + x_{\rho_{min}}(t_1) & (64)\end{aligned}$$

where t_1 represents the time at which the desired reference profile ρ_N^r (V_N^r) has been achieved by the cyber-attack system. The shock starts at a time $t_2 > t_1$, when $x_{\rho_{max}}(t_2) = x_{\rho_{min}}(t_2)$ is satisfied for some time t_2. Therefore we can find t_2 by equating (63) and (64), following which, the shock formation location $x_{\rho_{max}}(t_2)$ can be determined.

The other significant factor in choosing ρ_{min} and ρ_{max} is the velocity with which the shock travels, after it is formed. To calculate the shock speed, the generalized Rankine-Hugoniot shock condition is used, according to which, for a system represented in conservative form:

$$\frac{F(U)}{\partial t} + \frac{\partial G(U)}{\partial x} = 0 \qquad (65)$$

the shock, once formed, must necessarily satisfy the equation:

$$\tilde{\lambda}[F(U)] = [G(U)] \qquad (66)$$

where, $[X]$ denotes the magnitude of the discontinuous jump in the quantity X, and $\tilde{\lambda}$ is the velocity of propagation of the shock. Using (61) in (66), we arrive at:

$$\tilde{\lambda} = \frac{\rho_{max}V^{eq}(\rho_{max}) - \rho_{min}V^{eq}(\rho_{min})}{\rho_{max} - \rho_{min}} \qquad (67)$$

Using (67), a typical 3D-plot of the shock velocity as a function of ρ_{min} and shock magnitude $\Delta\rho$, where $\Delta\rho \equiv \rho_{max} - \rho_{min}$ is given in Fig 5. From this figure, it is evident that the shock velocity is a function of the shock magnitude $\Delta\rho$, and for around $\rho_{min} > 30$, the shock moves backward irrespective of its magnitude. A malicious attacker can make use of Figs 4 and 5 to determine the reference density/velocity profiles, as demonstrated in the next section.

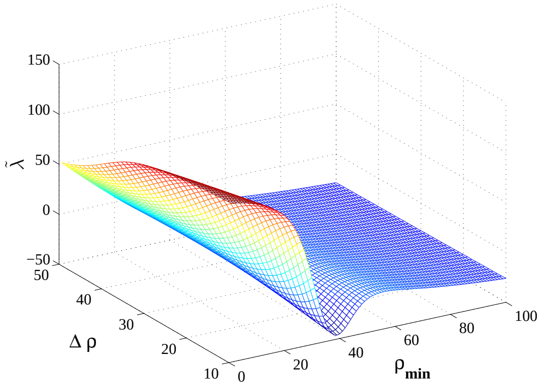

Fig. 5. Shock Velocity

VI. Numerical Results

A highway stretch of length $L = 10$ Km is considered, where the malicious vehicles intend to manipulate the shapes of the density/velocity profiles of the non-malicious vehicles. Two simulations are presented: their objectives being to manipulate the density profile assuming perfect information, and the velocity profile assuming imperfect information, respectively. The parameters used in these simulations for the model (7)-(10) are: $\tau = 15$ sec, $V_N^0 = 110$ kmph, $T = 1$ sec, $\gamma = 1$, $\rho_{max} = 160 \frac{\text{vehicles}}{\text{km}-\text{lane}}$, $A_0 = 0.008$, $\Delta A = 2.5A_0$. The boundary conditions are $\rho_N(t, 0) = \rho_N(0, 0)$, $V_N(t, 0) = V_N(0, 0)$, $\rho_M(t, 0) = \rho_M(0, 0)$. The simulations are performed in MATLAB by using the Lax Method [27] for discretization of the PDEs.

A. Density Profile Manipulation with Perfect Information

Assume that the attacker intends to create a shock of magnitude $\Delta\rho = 25 \frac{\text{vehicles}}{\text{km}-\text{lane}}$, and desires that the shock first forms at $x = L/2$ and then propagates backward on the highway. Following the arguments presented in Section V, from Fig 4, $\rho_{min} < 37.5$ and $\rho_{max} > 37.5$ are chosen. Then from Fig 5, by taking the plane corresponding to $\Delta\rho = 25$, it can be observed that the shock velocity is negative if $\rho_{min} > 25$. Hence a ρ_{min} is selected such that it satisfies $25 < \rho_{min} < 37.5$. Fig 6(a) shows the initial and reference spatial density profiles of the normal vehicles, with the latter profile chosen based on the above objectives. We point out that while the controller drives the density profile from its initial condition to the continuous reference profile of Fig 6(a), the strong solution of the PDEs is valid since the shock has not yet been formed. After this reference profile is attained, the traffic then spontaneously evolves into a shock, and after this shock forms, the weak solution of the PDEs is applicable.

Using the control law (28), this reference density profile is attained, as shown in Fig 7. Fig 8 shows the velocity changes performed by the malicious vehicles in order to achieve this objective. After the desired density profile is attained, it is assumed that the malicious vehicles exit the highway at around $t_1 = 8$ minutes. When this occurs, the space taken by the malicious vehicles is empty, thereby causing a decrease in the density of vehicles. This leads to an increase in the equilibrium velocity of the normal vehicles. This can be seen in Fig 12, wherein after the malicious vehicles have exited, the velocity of the normal vehicles V_N increases for the next 36 sec. For

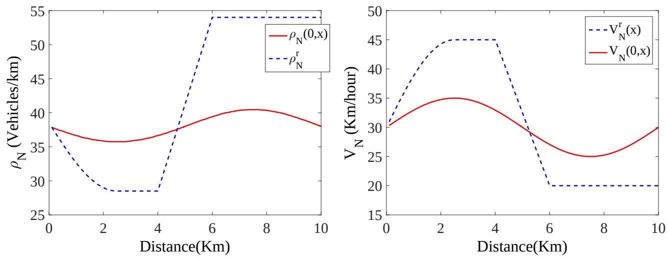

Fig. 6. (a) Initial and reference densities of normal vehicles: $\rho_N(0, x)$ and $\rho_N^r(x)$, (b) Initial and reference velocities of normal vehicles: $V_N(0, x)$ and $V_N^r(x)$

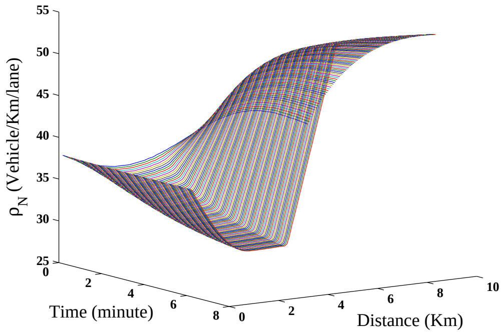

Fig. 7. Density of normal vehicles

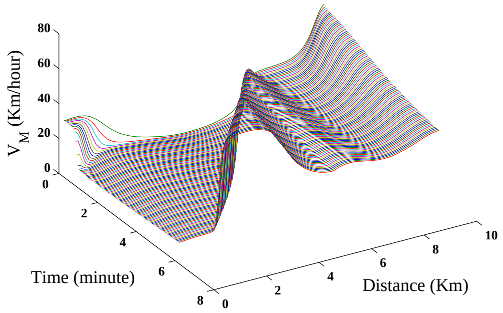

Fig. 8. Velocity of malicious vehicles

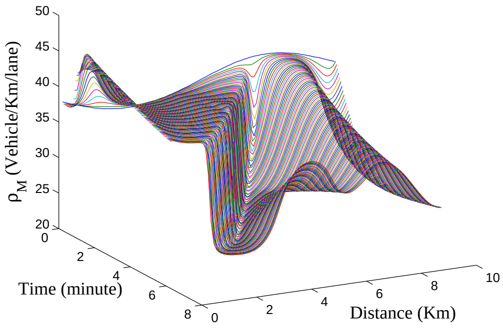

Fig. 9. Density of malicious vehicles

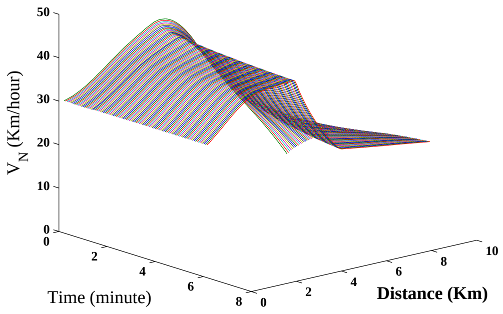

Fig. 10. Velocity of normal vehicles

$t > t_1$, even with the malicious vehicles no longer present, the intrinsic characteristics of the traffic cause this density gradient (as also the ensuing velocity gradient) to become progressively steeper and steeper, leading to a shock, which then propagates backwards along the highway. This is seen in Figs 11 and 12.

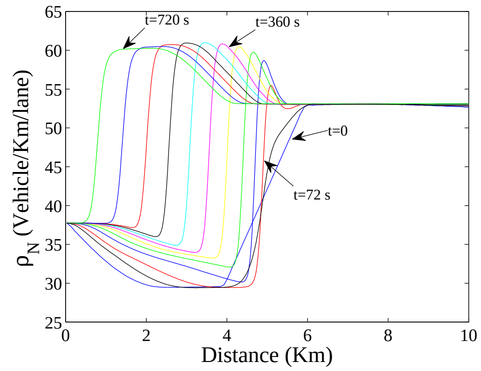

Fig. 11. Density of normal vehicles after exit (at $t_1 = 8 \; minutes$) of the malicious vehicles

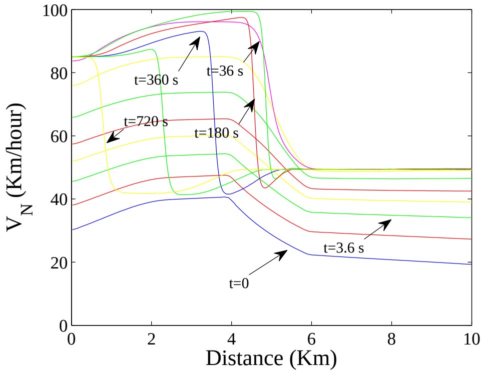

Fig. 12. Velocity of normal vehicles after exit (at $t_1 = 8 \; minutes$) of the malicious vehicles

B. Velocity Profile Manipulation with Imperfect Information

In this subsection, we show the effectiveness of the adaptive cyber-attack system derived in (36), (38), and (39) for imperfect information scenarios, wherein the attacker does not have access to exact values of T and τ. As before, the objective is to manipulate the velocity profile of the non-malicious vehicles in such a way that the profile automatically evolves into a shock wave that propagates along the highway. The initial and reference velocity profiles are shown in Fig 6(b). The initial values of $\hat{T}$ and $\hat{\tau}$ are 0.8 sec and 13 sec, respectively. The adaptation weights for h and w are $\gamma_1 = 10^{16}$ and $\gamma_2 = 10^3$, respectively, and the reasons for these values are as follows. It can be seen from Fig 17 that $\hat{\tau}$ and $\hat{T}$ have some oscillations before they reach their steady-state (and true) values of 15 sec and 1 sec, respectively. These oscillations cause the derivatives $\dot{\hat{\tau}}$ and $\dot{\hat{T}}$ to alternate between positive and negative values. Since γ_1 and γ_2 influence $\dot{\hat{\tau}}$ and $\dot{\hat{T}}$ (see (38) and (39)), it is important to choose them appropriately in order to ensure that

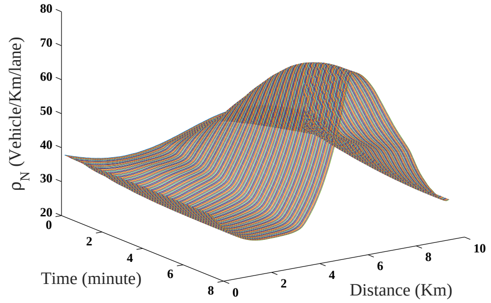

Fig. 13. Density of normal vehicles with adaptive cyber-attack system

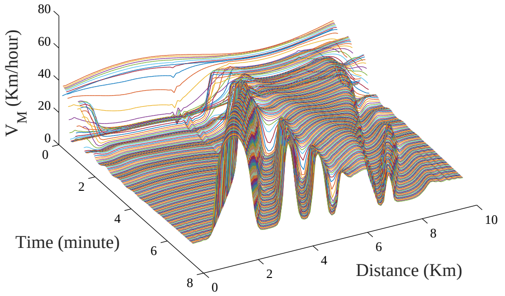

Fig. 14. Velocity of malicious vehicles with adaptive cyber-attack system

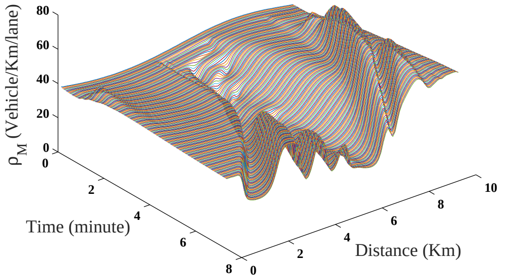

Fig. 15. Density of malicious vehicles with adaptive cyber-attack system

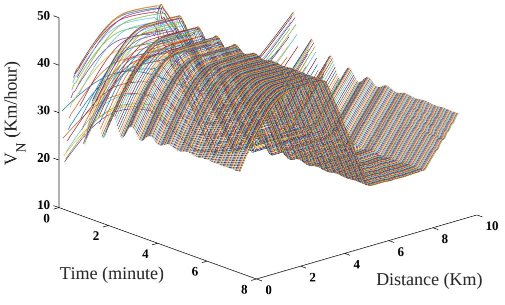

Fig. 16. Velocity of normal vehicles with adaptive cyber-attack system

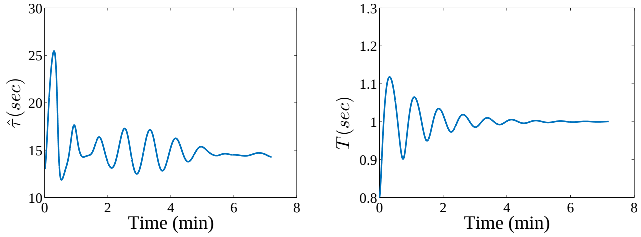

Fig. 17. (a) $\hat{\tau}$ and (b) $\hat{T}$ in adaptive cyber-attack system

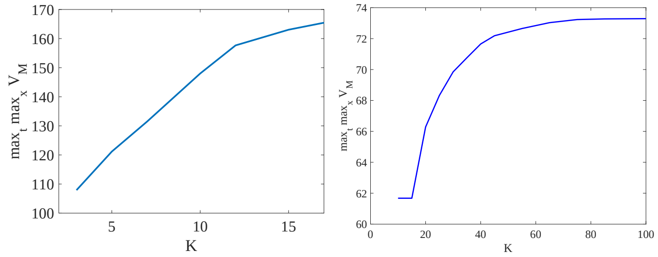

Fig. 18. $\max_t \max_x V_M$ vs. K in (a) Density profile manipulation and (b) Velocity profile manipulation

the estimates $\hat{\tau}$ and $\hat{T}$ remain positive at all times. As seen in Figs 13-16, the goal is attained, but with more oscillations in velocity and density. These oscillations occur because of the oscillations in $\hat{\tau}$, and $\hat{T}$ shown in Fig 17, which affect the decay rate of the Lyapunov fuction $\bar{Z}_V$ in (35). Finally, Fig 18 shows the effect of varying K_ρ and K_v on the maximum velocity of the malicious vehicles, required to achieve the density/velocity profile manipulation. As is evident, reducing K_ρ and K_v can make the attack more stealthy.

Remark: This study can lay a foundation for devising potential countermeasures to such attacks. The types of density and velocity wave profiles that could exist in traffic for the time interval before the shock is actually formed, are demonstrated. By monitoring the traffic for the onset of such density and velocity waves, warnings that such an attack is in progress can be triggered. Subsequently, personnel in the control room can initiate several countermeasures, such as modulating the traffic lights, giving advisories to the (normal) vehicles to adjust their speeds, and so on. Also, the analysis in this paper demonstrates one potential countermeasure which is to increase the relaxation time constant τ. When τ is large, a singular perturbation approximation is no longer valid for (60), and this physically means that the characteristic velocity will become different from that shown in Fig 4, thereby thwarting the intention of the attackers. Detailed development of such countermeasures is an avenue for future work.

VII. Conclusions

As autonomous vehicles become prevalent on highways, it is possible that an attacker may try to hack the driving software of some of the vehicles with malicious intent. The behaviour of an automated highway traffic system under the influence of such malicious agents, is analyzed using a two-species macroscopic model, with the two species being the malicious and the normal vehicles. The malicious vehicles are arbitrarily distributed among the normal vehicles, and seek to disrupt the traffic flow using subtle velocity changes that introduce undesirable wave effects in the traffic. Analytical control law expressions of the velocity changes of the malicious vehicles that generate a defined velocity/density profile on the highway, are determined for perfect and imperfect information scenarios. The specific case of the malicious vehicles generating a reference velocity/density profile that subsequently evolves into a shock, is demonstrated. This PDE-based analysis reveals the lack of resilience to the presence of malicious agents on automated highways, and calls for further research to develop suitable countermeasures.

Acknowledgements

The first two authors would like to acknowledge support from the National Science Foundation. This material is based upon work supported by the National Science Foundation under Grant IIS-1351677.

References

[1] Z. Guo, D. Shi, K. H. Johansson, L. Shi, “Optimal Linear Cyber-Attack on Remote State Estimation”, *IEEE Transactions on Control of Network Systems*, Vol. 4, No. 1, 2017.

[2] Y. Yan, Y. Qian, H. Sharif, “A Survey on Cyber Security for Smart Grid Communications”, *IEEE Communications Surveys & Tutorials*, Vol. 14, No. 4, 2012.

[3] P. Y. Chen, S. M. Cheng, K. C. Chen, “Smart attacks in smart grid communication networks”, *IEEE Communications Magazine*, Vol. 50, No. 8, 2012.

[4] E. Bou-Harb, C. Fachkha, M. Pourzandi, “Communication security for smart grid distribution networks”, *IEEE Communications Magazine*, Vol. 51, No. 1, 2013.

[5] E. S. Chang, A. K. Jain, D. M. Slade, S. L. Tsao, “Managing cyber security vulnerabilities in large networks”, *Bell Labs Technical Journal*, Vol. 4, No. 4, 1999.

[6] F. Khorrami, P. Krishnamurthy, Ramesh Karri, “Cybersecurity for Control Systems: A Process-Aware Perspective”, *IEEE Design & Test*, Vol. 33, No. 5, 2016.

[7] J. Reilly, S. Martin, M. Prayer, A. Bayen, “Creating complex congestion patterns via multi-objective optimal freeway traffic control with application to cyber-security”, *Transportation Research B*, Sep. 2016.

[8] I. Studnia, V. Nicomette, E. Alata, “Survey on security threats and protection mechanisms in embedded automotive networks”, *IEEE Conference on Dependable Systems and Networks Workshop*, 2013.

[9] M. Salfer, C. Eckert, “Attack surface and vulnerability assessment of automotive Electronic Control Units”, *12th International Joint Conference on e-Business and Telecommunications*, 2015.

[10] D. Ward, “Aligning safety and security systems for connected vehicles”, *Cyber-Security for Urban Transport Systems*, 2016.

[11] J. Wan, D. Zhang, S. Zhao, “Context-aware vehicular cyber-physical systems with cloud support: architecture, challenges, and solutions”, *IEEE Communications Magazine*, Vol. 52, No. 8, 2014.

[12] M. H. Lighthill, G. B. Whitham, “On kinematic waves II: A theory of traffic flow on long, crowded roads”, *Proc. Of the Royal Society of London Ser. A*, 1945.

[13] H. J. Payne, “Models of Freeway Traffic and Control”, *Simulation Council*, 1971.

[14] I. Prigogine and R. Herman, “Kinetic theory of vehicular traffic”, *Elsevier*, 1971.

[15] S. L. Paveri-Fontana, “On Boltzmann like treatments for traffic flow”, *Transportation Research Part B*, 1975.

[16] M. Treiber, A. Hennecke and D. Helbing, “Derivation, properties and simulation of a gas-kinetic-based, nonlocal traffic model”, *Physical Review E*, 59, 1999.

[17] S. P. Hoogendoorn and P. H. L. Bovy, “Continuum modeling of multiclass traffic flow”, *Transportation Research, Part B* 34, 2000.

[18] G. B. Whitham, “Linear and Nonlinear Waves”, 1974.

[19] W.F Phillips, “A kinetic model for traffic flow with continuum implication”, *Transportation Planning and Technology 5*, 1979, pp. 131-138.

[20] P.G. Michalopoulos, P. Yi, A.S. Lyrintzis, “Continuum modelling of traffic dynamics for congested freeways”, *Transportation Research B*, Vol. 27, No. 4, 1993, 315332.

[21] H. M. Zhang, “A Theory of Nonequilibrium Traffic Flow”, *Transportation Research B*, Vol. 32, No. 7, 1998, pp. 485-498.

[22] Swaroop D. and K. R. Rajagopal, “Intelligent Cruise Control Systems and Traffic Flow Stability”, *Transportation Research Part C*, *1999.*

[23] J. Yi, H. Lin, L. Alvarez and R. Horowitz, “Stability of Macroscopic Traffic Flow Modeling through Wavefront Expansion”, *Transportation Research Part B*, 2003.

[24] Y. Li, E. Canepa, C. Claudel, “Optimal Control of Scalar Conservation Laws Using Linear/Quadratic Programming: Application to Transportation Networks”, *IEEE Trans. on Control of Network Systems*, Vol. 1, No. 1, 2014.

[25] M. Ghanavati, A. Chakravarthy, P. P. Menon, “PDE-based Analysis of Automotive Cyber-Attacks on Highways”, *American Control Conference*, May 2017.

[26] H. Khalil, *Nonlinear Systems*, Prenctice Hall, 2002.

[27] R. J. Leveque, *Finite-Volume Methods for Hyperbolic Problems*, Cambridge Texts in Applied Mathematics, 2002.