A Framework for Broker Placement in Vehicular Ad hoc Networks

Laura Poff and Aniruddha Gokhale
Dept of EECS
Vanderbilt University
Nashville, TN 37235, USA
laurapoff@gmail.com, a.gokhale@vanderbilt.edu

Mark McDonald
Dept of Civil & Environmental Eng
Lipscomb University
Nashville, TN 37204, USA
mark.mcdonald@lipscomb.edu

***Abstract*—Vehicular networks are a classic example of mobile, peer-to-peer networks. The domain of intelligent transportation engineering (ITS) envisions a wide range of services over these vehicular networks. Supporting the different quality of service properties of these services requires effective dynamic resource management solutions in vehicular networks, which cannot be realized using vehicles alone due to factors such as varying vehicular mobility patterns, varying vehicular traffic density, and obstacles. Road side units (RSUs), which serve as event and data brokers among the moving vehicles, can alleviate this problem. To maximize the value proposition offered by RSUs requires their effective placement that holistically accounts for physical factors, such as traffic patterns and mobility, with cyber factors, such as communication protocols and messaging. To address these requirements, this paper presents a design methodology based on the principles of surrogate modeling wherein small-scale, micro-simulations of the ITS cyber-physical system are used to develop training points, which in turn are used to train a surrogate model. The surrogate model is subsequently used to make planning decisions for ITS.**

Keywords- VANETs; cyber-physical systems; capacity planning for resource management; modeling and design

I. Introduction

Intelligent Transportation Systems (ITS) are envisioned to address the numerous challenges faced by the transportation sector [1]. One category of solutions envisioned in ITS pertains to the real-time and reliable delivery of traffic-related information to drivers both for safety-critical applications (such as blind spot warnings during lane changing) and for applications that improve driving experience and help the environment (such as notification of congestion and rerouting advise that can help to alleviate traffic congestion and lost productivity).

The ITS services must be realized in the context of vehicular ad hoc networks (VANETs). Each of these services imposes different quality of service (QoS) demands that require effective dynamic resource management. However, resource management solutions are hard to realize using VANETs alone due to many factors, such as disparate traffic mobility patterns, traffic density, and obstacles to name a few. One way to alleviate these challenges is to use brokers that serve as mediators for event and data dissemination required in resource management solutions for the ITS services. In ITS, these brokers are called road side units (RSUs). An immediate question that arises is how to place these brokers that will maximize the efficiency of dynamic resource management solutions in VANETs.

Effective Broker placement in VANETs is a hard problem because decision techniques for broker placement must account for both the transportation-related challenges—*the physical dimension*, and the information technology challenges—*the cyber dimension* making it a cyber-physical system (CPS). It is neither economically feasible to deploy real infrastructure elements and test the effectiveness of these brokers for resource management, nor is it computationally feasible to simulate to conduct detailed simulations. Furthermore, a general lack of ITS-specific simulators, which can combine both the cyber and physical properties in the simulation, make the task even harder.

Therefore in this paper we present a framework for ITS broker (*i.e.*, RSU) placement decisions that can be made quickly and inexpensively, moving the design of the cyber-physical system closer to its real-world goal. To address the problems with scalability of simulations, we rely on using a surrogate model [2] which is trained using a relatively small number of training points obtained from a small set of micro-simulations that are quite inexpensive to execute. The chosen surrogate model is then used to make decisions regarding the system. The strength of this framework is such that once the system is characterized using a surrogate model, subsequently the model can be used in both the planning stages (*e.g.*, infrastructure decisions like placement of the RSUs) and at runtime once the system is built (*e.g.*, for dynamic resource management).

The rest of the paper is organized as follows: Section II discusses related research; Section III provides details of our methodology and the associated framework; Section IV demonstrates the application of our methodology to a sample scenario; and Section V discusses the implications of this framework and future work.

This work was supported in part by NSF CNS 0915976. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

II. RELATED RESEARCH

To the authors' knowledge the framework presented in this paper for decision-making has not been applied to any scale VANET problems. Parts of the framework we present, such as micro-simulation and optimization techniques, are in use within the field for infrastructure decisions as described below. Many optimization decisions in deployment of traditional event brokers also exist. A general observation is that these solutions either address the physical issues or the cyber issues in isolation, but not both at once.

Micro-simulation is suggested as the appropriate method for testing ITS services, over both Highway Capacity Manual (HCM) procedures and macroscopic simulation [3]. Both [3] and [4] provide suggestions on how best to build and validate realistic microscopic models for transportation systems. An actual ITS deployment is validated against micro-simulation models in [5], although the system is for adaptive traffic control systems and not for resource management in VANETs.

There are currently a sparse number of studies that compare micro-simulation results of VANETS to actual deployments as in [6], but there are studies that compare micro-simulation results to other mobility model [7], [8]. The general consensus, however, is that micro-simulation models can accurately represent real traffic systems when they are properly calibrated [3], [5], [9], [10]. In our case too since we cannot afford the computationally expensive large-scale micro-simulations, our methodology is based on surrogate modeling that leverages micro-simulations. Surrogate modeling is preferred over related techniques, such as response-surface models or even large-scale simulators and emulators because they are computationally less expensive compared to other techniques and are practical.

The use of optimization techniques for infrastructure deployment and system planning is also already in use within the ITS domain. Examples include maximizing coverage while guaranteeing a minimum amount of coverage time [11]; maximizing the reliability of information dissemination [12]; bandwidth minimization as well as travel-time minimization [13]; and maximizing the utility that comes from hardware distribution and information gathering [14].

Many prior works exist in the cyber realm on modeling and capacity planning. A recent work [15] develops analytical models for mission-critical publish/subscribe systems that can help to answer a number of questions, such as whether a certain deployment topology is effective or not. Although the authors account for multiple layers of the networking stack including the physical level, this research is still confined to the cyber world. The CCD [16] is a content-based pub/sub middleware that provides solutions to optimally place operators within brokers for customized content delivery to subscribers. This work focuses on placing appropriate functionality (*i.e.*, operators) within brokers of a pub/sub environment but not the placement of brokers.

Our research proposes extending current decision-making practices in ITS development and deployment to include surrogate models in order to characterize the system more efficiently than just micro-simulation and optimization can do alone. In other areas of engineering design, the process of using training points found through simulation to train a surrogate model has been demonstrated, *e.g.*, Structural Engineering [17], Pavement Design [18] and Transportation Engineering [19] provide a few examples where this framework has been utilized.

III. A METHODOLOGY FOR CAPACITY PLANNING IN INTELLIGENT TRANSPORTATION SYSTEMS

We now describe the details of our design-time methodology for placement of event brokers, such as RSUs, in VANETs to support effective resource management. Our methodology accounts for both the cyber- and physical-level challenges.

A. Overview of the Capacity Planning Methodology

Figure 1 illustrates the four steps of our methodology, which are described in detail in the rest of this section.

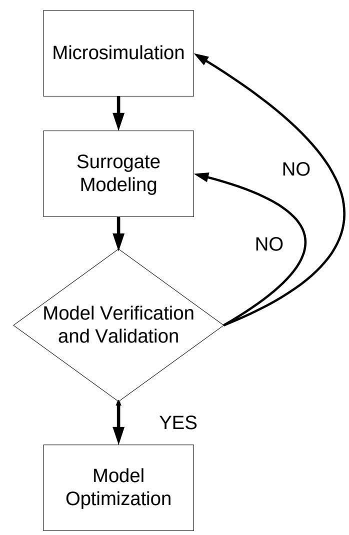

Figure 1: Methodology

1. **Collecting training points** – A small number of micro-simulations involving the CPS properties of VANETs are conducted to collect training points needed to develop the surrogate models for VANETs.
2. **Building surrogate models** – the second step is to develop the surrogate models using the training points for the system under study. We use the Gaussian Process modeling approach to build these surrogate models.
3. **Validating the surrogate models** – once the surrogate models are developed, they must be validated for accuracy and correctness, which forms our third step.

4. **Optimizing the decisions** – once the models are validated, we use optimization techniques to make optimized engineering decisions.

Surrogate modeling provides an inexpensive but highly accurate design-time solution for designing complex systems. We believe that such a methodology has not been exploited in VANETs (and ITS in general) though it is popular in traditional engineering designs.

B. Microsimulation to Collect Training Points

Training points are needed to build surrogate models. The challenge lies in how the training points can capture both the physical and cyber properties of VANETs all at once unlike traditional engineering designs that focus only on the physical properties. Thus, in our experimentation, we have chosen decision variables (*i.e.* inputs) for the physical dimension, such as vehicular speed and density, and wireless power beacon. The decision variables we considered on the cyber side included protocols used and data packet sizes. The output parameters of the training should ideally provide some measure of the quality of the system. For VANETs, the output should describe how well communication is being conducted throughout the system (*e.g.*, number of successfully transmitted packages and available communication window for a moving vehicle when communicating with a RSU). The example presented here utilizes speed of a vehicle moving by an RSU and power beacon level of the RSU.

The entire domain of the possible region for inputs must be covered when collecting the training points. This is easy to visualize for one or two dimensional problems (with 1 or 2 inputs) but can become difficult to verify as dimensionality increases. In the case of higher number of dimensions, random sampling and importance sampling are promising techniques to fulfill this requirement. Due to the iterative nature of our approach, refinements to these training points are likely as explained in Section III.D where the Predicted Residual Sum of Squares (PRESS) test is used to choose the most appropriate model.

C. Surrogate Modeling: Gaussian Process

In the second step, the collected training points are used to train a surrogate model, which is a mathematical model that provides a mapping from the input values to the output values while obscuring the physics of the system. Currently there are several surrogate modeling approaches that are considered appropriate to use in a variety of design and decision making capacities; some of these are conventional response surface models, polynomial chaos, Gaussian Process (GP), and radial basis functions [2].

Among the many different techniques available for surrogate modeling, we have chosen Gaussian Process (GP) models due to the flexibility they provide [2]: (1) GPs require no specific functional form as in regression-based techniques; all systems that are smooth and continuous can be modeled. Even the discontinuous nature of wireless systems can be captured and modeled in the continuous field of a GP due to the aggregation level of the inputs and outputs; (2) The trend function of the input variables can be varied and compared in order to find the best possible fit; (3) The GP can be refined using the locally defined variance function and minimization techniques; and (4) GP can also handle a large number of inputs (30-50) so it is able to model complex configurations of VANETs.

The process of training a GP-based surrogate model uses the training values, a correlation function between the training points, and a trend function in order to build a likelihood function. Parameters are then estimated by maximizing the likelihood. The GP with its estimated parameters can then be used to find *E[output]* and *E[Var(output)]* for any input parameters. The process utilized in this research is described below. Note that capitalization indicates that the value is a vector or matrix, while lower case letters refer to a scalar.

1. Put the training values (output) in an *m-by-1* vector, Y. Enter the normalized training points (input parameters) in an *m-by-n* matrix, where *m* is the number of training points and *n* is the number of input dimensions (21-by-2 in our case shown in Section III.D).

2. Build the correlation matrix between the training points, R (*m-by-m*). R can take many forms, but the Gaussian form is considered in this research, which is built by assigning each element, c, according to the following equation.

$$c(x^j, x^k) = \exp\left[-\sum_{i=1}^{n} \xi_i (x_i^j - x_i^k)^2\right] \qquad (1)$$

   Where ξ_i is a scale factor that is estimated in step 4 below and x_i^j and x_i^k represent the j^{th} and k^{th} training points at the i^{th} dimension. Note that in practice, it is easiest to solve the minimization in step 4 by searching on an exponential scale. In Equation 1, ξ_i becomes $\exp(\psi)$ and the minimization in Equation 4 is done in terms of ψ, β, and σ. R is used to make the covariance function using the process variance σ^2 that is estimated in step 4.

$$Cov() = \sigma^2 R \qquad (2)$$

3. Build the trend function, F *(m-by-depends on order)* according to the mean function found below. The trend function can help capture any underlying trends in the input variables.

$$\mu(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \dots \qquad (3)$$

   Where x_1 is the speed (mph) and x_2 is the power (mW). If a zero order GP is chosen, then the mean function only consists of the first term, whereas if a first order GP is chosen three terms are included. F is formed so that F $\beta = \mu$. In a zero order GP, F is a vector of ones *(m-by-1)*; in a first order GP, F is a vector of ones augmented with the matrix of normalized training points *(m-by- n+1)*; in a second order GP, F is further augmented with the squared values of the normalized training points *(m-by- 2n+1)*.

4. Find the optimum correlation parameters by finding the minimum of the negative log likelihood given by the following equation.

$$\min_{\xi,\beta,\sigma^2} \; -\ln[L(\cdot)] = m\ln(\sigma^2) + \ln|R| + \frac{(Y-\mu)^T R^{-1}(Y-\mu)}{\sigma^2} \quad (4)$$

Equation 4 in its entirety is not used in practice since it can become very computationally expensive. Reference [2] recommends using the *concentrated ln-likelihood function* that consists of the first two terms of Equation 4 (shown in Equation 5). Then the maximum likelihood estimates of β and σ^2 can be computed in equations 6 and 7.

$$\min_{\xi,\beta,\sigma} \; -\ln[L(\cdot)] \approx m\ln(\sigma^2) + \ln|R| \quad (5)$$

$$\beta = (F^T R^{-1} F)^{-1}(F^T R^{-1} F) \quad (6)$$

$$\sigma^2 = \frac{1}{m}(Y-\mu)^T R^{-1}(Y-\mu) \quad (7)$$

5. Use the parameters to calculate the estimated value of new points, equation 8. One of the benefits of the GP is that the variance of the new points can also be computed due to the covariance function and correlation matrix, equation 9. Computing the variance can help in any type system design since the variance can be minimized as part of an optimization statement if that is desired.

$$E[y(x^*)|Y] = f\beta + r^T R^{-1}(Y-\mu) \quad (8)$$

$$Var[y(x^*)|Y] = \sigma^2(1 - r^T R^{-1} r) \quad (9)$$

Where r is the correlation vector of the new point x* with the training points and f is the trend vector formed with the new point.

D. Model Selection and Validation

The next step requires us to select the right model and validate it for accuracy. We consider three different underlying trend functions and choose the model that minimizes the error between the predicted values and the simulated values. When choosing from a variety of possible surrogate models (*e.g.*, zero, first, or second order GP as used in our research) our methodology leverages the Predicted Residual Sum of Squares (PRESS) test [20], which is chosen because it is a computationally less expensive cross validation technique. The steps in PRESS are summarized below:

1. Individually omit each *i* th observation; recalculate the fitted model for remaining *n-1* data
2. Calculate the prediction error for the *i* th observation and square the difference
3. Repeat the process for all *n* observations and compute the sum of squares
4. Compare sum of squares value to other candidate models, with lowest value preferred

The model that is chosen must also be validated. A likelihood ratio metric is used, which involves developing two competing hypotheses and the ratio of the likelihood of observing the validation data (new points) conditioned on each hypothesis (found below) is computed. The likelihoods are computed by finding the product of the probabilities for each new point given the hypothesis. For our methodology, a Bayes validation metric is used. The Bayes method involves choosing two competing hypotheses. Ideally they should be mutually exclusive and collectively exhaustive. The Bayes method provides a flexible framework for drawing inferences, which is the reason we chose it in our methodology.

The Bayes factor (B) for the competing hypotheses is computed. A Bayes factor is the ratio of the likelihood of observing the validation data (new points) conditioned on each hypothesis. The likelihoods are computed by finding the product of the probabilities for each new point given the hypothesis. If B is greater than one, the model described by H_1 (The zero order GP model) is supported over the model (H_2) that is not described by equation 10 below.

$$B = \frac{L(observation|H_1)}{L(observation|H_2)} = \frac{\prod_{i=1}^{p} P(observation_i|H_1)}{\prod_{i=1}^{p} P(observation_i|H_2)} \quad (10)$$

where H_1 and H_2 are described below:

H_1: The simulation output is normally distributed and has a conditional mean and variance given by the zero order GP model.

H_2: The simulation output is normally distributed and is not conditioned on the zero order GP model; the mean and variance are given by the observed values (*i.e.*, training points).

E. Optimization of Engineering Decisions

The final step in our process is that of optimizing the chosen surrogate model, which can then be used to make design and real-time decisions regarding engineering systems. An optimization statement that utilizes a surrogate model can either use the surrogate model as part of the objective statement or within the constraints. For there often are multiple objectives that need to be maintained -- some regarding the physical possibilities and limitations of the system and some regarding the cyber aspects. This framework utilizes the method of reducing multiple objectives to a single statement with a weighting scheme [2], [21]. The surrogate model output is then used as a constraint that sets some minimum expectation of quality for the system.

A generalized statement can be found below in equation 11.

$$\begin{aligned}&\text{minimize} \sum_{i=1}^{k}(\mathrm{w}_i)[F_i(x)]^p \\ &\text{such that} \\ &q \geq q_{\min} \\ &x_{\min} \leq x \leq x_{\max}\end{aligned} \tag{11}$$

Where $F_i(x)$ represents the various objectives that can be positive when F_i should be minimized and negative when F_i should be maximized; $w_i>0$ describes the relative weights given to each design variable such that $\Sigma w_i=1$; $p>0$ higher values of p can more effectively lead to a full Pareto front [21]; q describes some quality of service measure (the surrogate model output for this framework) and q_{min} describes the minimum acceptable value of q; and the last two constraints represent constraints that ensure the design variables remain within realistic bounds.

IV. Demonstrating the Methodology on a Resource Management Use Case

The example we provide demonstrates the use of the proposed methodology using simulation results previously done in [22]. This research sends a car through the communication range of an RSU in order to test the quality of the information under different system configurations using the 802.11p standard. The speed of the vehicle and level of the power beacon are varied in an attempt to characterize the performance of the system using the length of the communication window. We re-examine the system under the new methodology and provide a more thorough characterization of the system, which can then be systematically used in decision-making. This methodology addresses the need that cyber-physical system research has regarding a new and more integrated approach to system design and functionality [23].

A. Obtaining Training Points

In this example, inputs were chosen as speed and level of the power beacon. The output of the model is the length of the communication window. Speed and power both have significant relationships with the communication window, which is a key parameter in making the right resource management decisions. For example, based on the QoS needs of the application, power level can be increased or decreased, or speed can be slowed or increased. Note that this is just one of the many possibilities, which forms our future work.

The configuration of the VANET is set up in OMNeT++, which is an open source software using the 802.11p protocol. The RSU sends packets of information to the vehicle. The information that is logged from this simulation is the time at which the RSU finds, authenticates, associates with, and loses the vehicle. The communication window is defined as the difference in the time the RSU loses the vehicle and the time at which the RSU associates with the vehicle and constitutes all of the conceivable time that the vehicle and RSU can communicate with each other.

In initial simulations of this system performed in [22] the power of the beacon was held steady at 5mW while the speed of the vehicle varied from 15 mph to 70 mph using 11 discrete speeds. The speed was then held constant at 60 mph while the power beacon was varied from 5 mW to 50 mW (shown in Figure 2). The communication window results were intuitive in that as power beacon increased while speed was held constant, the communication window also increased. Speed and communication window had an inverse relationship, as expected.

While these original test locations give some insight into the cyber-physical system, they are not ideal for the next step in the framework: building a surrogate model. In order to trust the surrogate model, the training data should cover the entire domain. New simulations were run using training points that covered the desired prediction area. Figure 3 shows the location of the newly selected points.

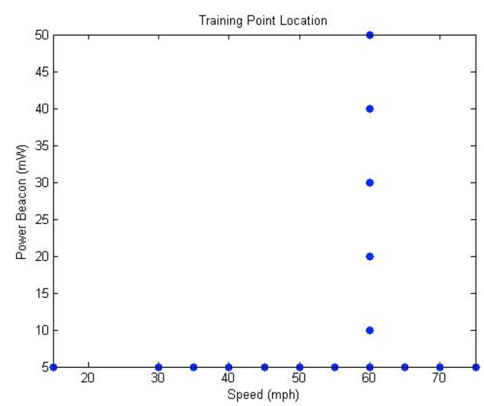

Figure 2. Original Data

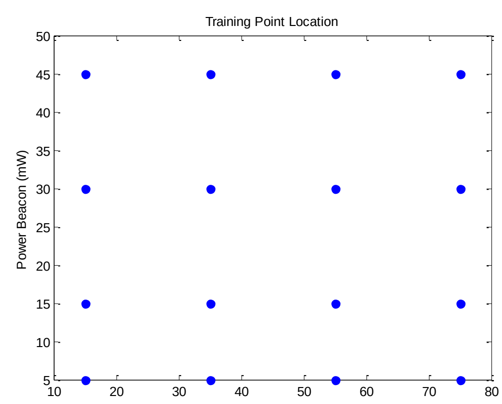

Figure 3. New potential training points

As this framework progresses, however, more training points were added in strategic locations to decrease the error in the competing models. These new training point locations (shown in Figure 4) represent the parts of the system that undergo the most change in the communication window with respect to changes in power and/or speed. This process is explained in Section III.D where the PRESS test is used to choose the most appropriate model.

The communication window results from the complete set of training data are shown below as a set of three tuples[1] sorted by the length of the communication window. Each tuple is arranged as input values (speed in mph and power in mW)– the training points, and the output value (communication window in seconds) – the training values. The final training points and values are:

```
<75, 5, 9.45>, <55, 5, 12.78>, <75, 15, 16.47>, <35, 5, 20.36>, <55, 15, 22.27>, <75, 30, 23.57>, <75, 45, 28.57>, <70, 40, 28.77>, <55, 30, 31.87>, <35, 15, 35.17>, <55, 45, 39.07>, <25, 10, 40.37>, <15, 5, 47.82>, <35, 30, 50.47>, <35m 45, 61.47>, <25, 40, 81.17>, <15, 15, 83.47>, <25, 45, 86.07>, <15, 30, 117.57>, <15, 40, 136.37>, <15, 45, 144.67>
```

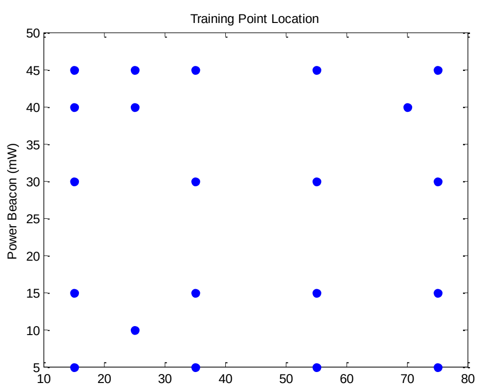

Figure 4. Complete and Final set of training points

B. Building, Selecting and Validating a Surrogate Model

We used Matlab to build each Gaussian Process model using several m-files originally written and used in [17]. Both the PRESS test and the Bayes validation method is demonstrated in this section.

1) Model Selection

Originally the PRESS values computed when using the training points found in Figure 3 were high, as seen in Table I. Therefore, an analysis of the error produced by omitting each point during the PRESS test was performed. This analysis is shown in Figure 5.

The three numbers next to each point represent the squared error that arises from removing that point from the set of training data while training the zero, first, and second order GP models respectively and then trying to predict the point. The points that were not predicted well while using any of the three models (*i.e.* contributed a value greater than 100 to the entire PRESS value) are circled in Figure 5. The behavior of the system around these areas (note that the areas are near the edges of the GP) is harder to predict from the remaining training points after the point in that each area is removed. In order to better predict these points, training points closer to them are added, resulting in the full set of training points presented earlier in Figure 4. This provides lower PRESS values for all models as seen in Table I.

Table I. PRESS values for GP models

Order of trend function	**0**	**1**	**2**
Figure 3 training points	655.48	868.03	1143.56
Figure 4 training points	225.75	302.47	375.88

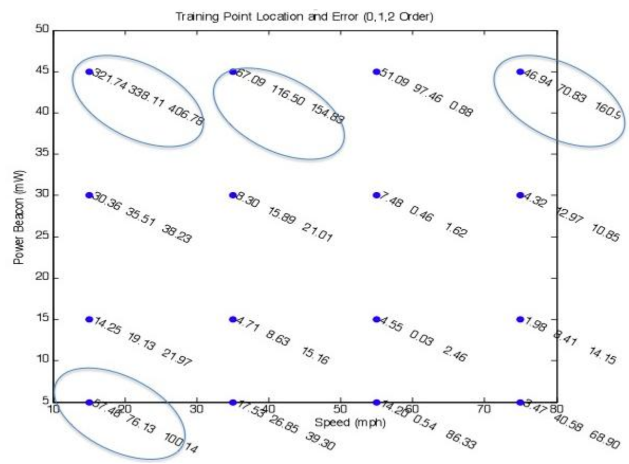

Figure 5. PRESS value resulting from each point (0,1,2 order)

Both sets of training data support the zero order GP, which is what was adopted for this research; the model is shown in Figure 6 and the contours describe the length of the communication window with respect to the Power Beacon and Speed. The estimated values for the parameters are: $\exp(\psi)=[-0.421, -1.987]$ or $\xi=[0.656, 0.137]$; $\beta=38.412$; and $\sigma^2 = 4121.990$.

2) Model Validation using Bayes Factor

Before the chosen model is used, it must be validated. A Bayesian form of validation is used in this research as explained earlier. The GP predictions for five new arbitrarily chosen points and all other information necessary to finding the Bayes factor[2] resulted in a factor of about 588,270, which supports the hypothesis that our data is normally distributed conditioned on the model, as opposed to being normally distributed not conditioned on the model. Therefore, the zero order GP is retained and used to make optimized decisions regarding the VANETs.

[1] Due to space constraints the data could not be presented in tabular form.

[2] Bayes validation table not presented due to space constraints

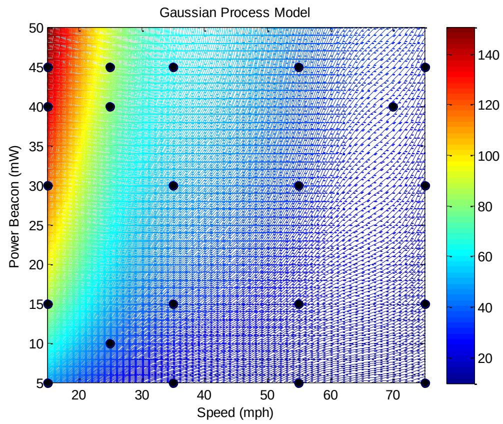

Figure 6. Zero order GP model

C. Optimizing the Decisions

In this example, the decisions that need to be made involve the speed of the vehicle and the strength of the power beacon such that a certain communication window is maintained in the system. It would be unrealistic to assume that policy makers would significantly lower speed limits in order to meet communication window lengths, so part of the optimization statement in this example maximizes the speed, while still keeping it within the accepted speed limit. In a VANET, the power of the beacon affects cost in both power consumption and the necessity to recharge the battery. In order to preserve energy and cost, the optimization statement minimizes the strength of the power beacon. This research sets the communication window that is desired as 15 seconds and performed the optimization using the statement below.

$$\begin{aligned}
&\text{minimize } w_2(x_2^*) - w_1(x_1^*) \\
&\text{such that} \\
&y(x_1 x_2) \geq 15 \text{ (s)} \\
&45 \leq x_2 \leq 75 \text{ (mph)} \\
&1 \leq x_1 \leq 50 \text{ (mW)}
\end{aligned}$$

Where x_1 is the speed (mph) and x_2 is the power (mW), and the asterisk on both indicates that the values have been normalized in the minimization statement; $y(x_1, x_2)$ is the communication window that is predicted using the zero order GP model with the potential new input values, $[x_1, x_2]$.

The first constraint requires using the GP model to compute the communication window for potential input values, which requires computing the correlation vector, *r*, between the new potential point and the training points. Then $E[y(x^*)|Y]$ is computed as described in Section III.C step 5. If the value computed for the communication window is less than 15 seconds, then a very high value is returned for the value of the objective function in order to deter a solution at that point.

The previous result assumes that the decisions regarding speed and power in this system are weighted equally (or are equally important). The optimization statement could easily be altered in order to weight one of the input parameters over the other. This is dependent on the resource management decision. Therefore, the weight of each parameter is varied from 0 to 1 such that the sum of the weights is one. In Figure 7, the weight of the speed term is plotted versus the solution found for speed to see the trade-off between the two.

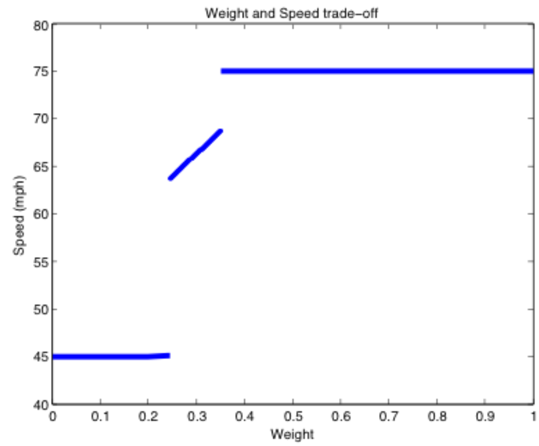

Figure 7. Weight and speed trade-off

The results are intuitive, when the weight in front of the speed term is higher than the solution results in a higher speed as expected from the objective to maximize speed. In order to effect a change in the optimal speed and power allocation of the system, the speed must be weighted at close to 0.35. The outcome would also be different if the constraints were relaxed on the second constraint regarding speed.

This demonstration has yielded optimal system parameters for the system at hand. In general, the results show that this framework can be used when developing an ITS. The framework can be used for the deployment of a full scale ITS when input and output variables are chosen carefully. The framework has been described in a way that also allows the surrogate model to be used when the system is deployed for real-time computation and information dissemination.

V. Conclusions

In building a model for solving resource management problems in VANETs using brokers, this research has extended the surrogate modeling technique used in traditional engineering design to characterize a cyber-physical system. Our preliminary results have yielded optimal system parameters for the system at hand. The results show significant promise for the framework in developing ITS services. The framework can be used for the deployment of a full scale ITS services in VANETs when input and output variables are chosen carefully.

Several challenges were encountered and lessons learned when building the surrogate model. The location of training points must be perfected for systems. Future work will include methods of adaptive sampling in order to minimize the number of training points needed. Also, the PRESS test results seemed too high for most of the competing models. Perhaps better training points would help to reduce the PRESS test results. A wider range of surrogate models could be tested for better possible models. The GP model that was ultimately adopted is very intuitive, but so were the ones that were rejected in that they all show an inverse relationship between speed and communication window and a direct relationship between power beacon and communication window.

A Bayes hypothesis test was performed that supported choosing the zero order GP with conditional mean and variance over a normal distribution with a mean and standard deviation that came from the observed training values. In the future a complete error analysis should be performed in order to guarantee that competing errors are not cancelling out, leading to a high Bayes factor, and incorrectly leading us to adopt the model.

Our future work is identifying ways to generalize the framework to solve multiple different problems in collaborative, peer-to-peer networks.

REFERENCES

[1] M. Bowman, S. K. Debray, and L. L. Peterson, "Reasoning about naming systems," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 15, no. 5, pp. 795-825, 1993.

[2] A. I. J. Forrester, A. Sobester, A. J. Keane, and A. I. of A. and Astronautics, *Engineering design via surrogate modelling: a practical guide*. Wiley, 2008.

[3] R. Dowling, A. Skabardonis, and V. Alexiadis, "Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software," FHWA, U.S. Deparment of Transportation, FHWA-HRT-04-040, 2004.

[4] E. Brockfeld, R. D. Kühne, A. Skabardonis, and P. Wagner, "Toward benchmarking of microscopic traffic flow models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1852, no. 1, pp. 124-129, 2003.

[5] C. Kergaye, A. Stevanovic, and P. T. Martin, "Comparative Evaluation of Adaptive Traffic Control System Assessments Through Field and Microsimulation," *Journal of Intelligent Transportation Systems*, vol. 14, no. 2, pp. 109-124, 2010.

[6] E. Alba, S. Luna, and J. Toutouh, "Accuracy and efficiency in simulating vanets," *Modelling, Computation and Optimization in Information Systems and Management Sciences*, pp. 568-578, 2008.

[7] J. Harri, F. Filali, and C. Bonnet, "Mobility models for vehicular ad hoc networks: a survey and taxonomy," *Communications Surveys & Tutorials, IEEE*, vol. 11, no. 4, pp. 19-41, 2009.

[8] C. Sommer and F. Dressler, "Progressing toward realistic mobility models in VANET simulations," *Communications Magazine, IEEE*, vol. 46, no. 11, pp. 132-137, 2008.

[9] R. Dowling, A. Skabardonis, J. Halkias, G. McHale, and G. Zammit, "Guidelines for calibration of microsimulation models: Framework and applications," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1876, no. 1, pp. 1-9, 2004.

[10] J. B. Lee and K. Ozbay, "New Calibration Methodology for Microscopic Traffic Simulation Using Enhanced Simultaneous Perturbation Stochastic Approximation Approach," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2124, no. 1, pp. 233-240, 2009.

[11] O. Trullols, M. Fiore, C. Casetti, C. F. Chiasserini, and J. M. Barcelo Ordinas, "Planning roadside infrastructure for information dissemination in intelligent transportation systems," *Computer Communications*, vol. 33, no. 4, pp. 432-442, 2010.

[12] M. Sardari, F. Hendessi, and F. Fekri, "Infocast: a new paradigm for collaborative content distribution from roadside units to vehicular networks," in *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON'09. 6th Annual IEEE Communications Society Conference on*, 2009, pp. 1-9.

[13] C. Lochert, B. Scheuermann, C. Wewetzer, A. Luebke, and M. Mauve, "Data aggregation and roadside unit placement for a vanet traffic information system," in *Proceedings of the fifth ACM international workshop on VehiculAr Inter-NETworking*, 2008, pp. 58-65.

[14] T. G. Crainic, M. Gendreau, and J. Y. Potvin, "Intelligent freight-transportation systems: Assessment and the contribution of operations research," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 6, pp. 541-557, 2009.

[15] A. Schröter, G. Mühl, S. Kounev, H. Parzyjegla, and J. Richling, "Stochastic performance analysis and capacity planning of publish/subscribe systems," in *Proceedings of the Fourth ACM International Conference on Distributed Event-Based Systems*, 2010, pp. 258-269.

[16] H. Jafarpour, B. Hore, S. Mehrotra, and N. Venkatasubramanian, "CCD: efficient customized content dissemination in distributed publish/subscribe," *Middleware 2009*, pp. 62-82, 2009.

[17] B. J. Bichon, "Efficient Surrogate Modeling for Reliability Analysis and Design," Vanderbilt University, 2010.

[18] J. Retherford and M. McDonald, "Estimation and Validation of Gaussian Process Surrogate Models for MEPDG-Based Sensitivity Analysis and Design Optimization," Washington D.C., 2011.

[19] S. Pathak, M. McDonald, and S. Mahadevan, "A framework for designing policies for networked systems with uncertainty," *Decision Support Systems*, vol. 49, no. 2, pp. 121-131, May 2010.

[20] D. M. Allen, "The Relationship between Variable Selection and Data Agumentation and a Method for Prediction," *Technometrics*, vol. 16, no. 1, pp. 125-127, Feb. 1974.

[21] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and multidisciplinary optimization*, vol. 26, no. 6, pp. 369-395, 2004.

[22] A. Gokhale, M. McDonald, S. Drager, and W. McKeever, "A Cyber Physical Systems Perspective on the Real-time and Reliable Dissemination of Information in Intelligent Transportation Systems," *Network Protocols and Algorithms*, vol. 2, no. 3, pp. 116-136, 2010.

[23] B. H. Krogh et al., *Cyber-Physical Systems, Executive Summary*. CPS Steering Group, Washington DC, March, 2008.