A SVM-based Approach for VANET-based Automatic Incident Detection

Luyi Chen, Ping Xu, Tiaojuan Ren, Yourong Chen, Binbin Zhou*, Hexin Lv*

College of Information and Science Technology, Zhejiang Shuren University
Hangzhou, Zhejiang, China

***Abstract* — Traffic incidents detection has been a crucial problem in the past decades, due to the considerable economical cost and inestimable disgruntlement from numerous drivers. In this paper, we have presented a SVM-based approach for Automatic Incident Detection (AID), in which the traffic data are collected by vehicular ad hoc networks (VANET) techniques. We have processed collected data and utilize traffic variables in the SVM model to confirm whether an incident occurs. Several experiments have been conducted to evaluate our approach's performance, and the results show that our approach could outperform the other two approaches in most cases.**

Keywords - automatic incident detection (AID), support vector machine (SVM), machine learning (ML)

I. INTRODUCTION

Traffic congestion has become a huge and increasingly severe problem worldwide nowadays, due to the growing demand on transportation and constraint resources supported by existing traffic infrastructures. Traffic incidents play a crucial role in the traffic congestion problem. In this way, incidents refer to abnormal events occur to obstruct the normal satiny traffic flows and affect the utilization of traffic infrastructures, i.e. traffic accidents, interception because of hazard weather conditions. Hence, automatic incident detection (AID) has been proposed and developed in the past decades, and attracted the interest of a number of scholars. Accurate and effective incidents detection could be helpful not only to relieve congestion, improve traffic efficiency and decrease fuel cost, but also provide reliable information to drivers to reduce their travelling time. Usually, a large amount traffic data utilized for AID has been a sharp problem. Data collection methods using current detectors (i.e. inductive loops and video cameras) have lots of shortcoming, e.g. the limited detection range and high costs of implementation and maintenance. Hence, we employ sensors nodes, which are widely used in vehicular ad hoc networks (VANET), to detect, transmit and fuse traffic data.

There has been a large amount studies adopting various techniques for AID in recent years [1]-[15]. Previous research generally can be categorized into four groups, machine learning (ML)-inspired algorithms, time series analysis (TSA), other comparative approaches and hybrid approaches. ML-based methods focus on traffic patterns and estimate current detected traffic variables whether it's incident-free [2]–[7]. TSA approaches underline dynamic and abnormal changes of traffic [8]–[11]. There are also some comparative approaches [12][13] and hybrid approaches[14][15]

In this paper, we have employed a support vector machines (SVM)-based approach to detect the VANET-based automatic incidents. SVM can be used for data analysis and pattern recognition through its supervised learning models companied with associated learning algorithms [16] [17]. SVM are effective tools in a broad area of classification problems and robust to irrelevant features [18]. We have extracted the most critical features related to incidents occurrence, such as speed, occupancy and volume. Furthermore, we have trained SVM through various features combinations. Finally, we have conducted experiments to evaluate the proposed approach's performance on a publicly available dataset containing real-world traffic data in California, which is used in a wide range of relevant studies. The simulation results presents that our approach can outperform relevant state-of-the-art approaches in three well-acknowledged evaluation metrics.

The remainder of this paper is organized as follows. We review some most representative relevant research in Section II, and describe the problem formulation in Section III. Section IV explains our SVM-based mechanism in details. Section V elaborates the design of experiments and results. Finally, we conclude the paper and outlines future research directions in Section VI.

II. LITERATURE REVIEW

In the past decades, there has been a large amount of research efforts paid on AID, and we would review the more representative relevant works using various AID techniques categorized.

A. *Application Fields*

Previous AID approaches are mainly applied into two fields, freeways and urban roads. These two application areas have different traffic characteristics. In freeways, traffic flow would present in a smooth and satiny way with various traffic density, which result in relatively homogenous traffic patterns [19]. On the contrary, traffic flow in urban zones are guided and controlled by traffic signals and traffic police, which would lead to a remarkable difference of traffic pattern compared with in freeways.

B. Detection Techniques

Various techniques applied in AID can be classified into four groups, machine learning (ML)-inspired algorithms, time series analysis (TSA), other comparative approaches and hybrid approaches.

Machine learning-inspired approaches usually treat AID as a binary classification problem, and explore traffic patterns in order to group the detected and processed data into one pattern to decide whether it's incident-free or not. Neural networks (NNs) have been successfully applied in AID in freeways in the late 1990s [20]-[22]. A constructive probabilistic NN (CPNN) has been employed in freeways for incidents detection [2]. A modular architecture for NNs has been proposed with traffic data collected by inductive loop detectors to monitor various transportation problems on signalized urban arterials [23]. The Srinivasana group proposed a reduced multivariate polynomial-based NN for incidents detection in freeways [3]. The application of SVM-based methods in AID has become popular in the early 2000s [4], [5]. Cheu et al. [4] applied several SVM in arterial network, and compared the performance with MLD and PNN in some metrics, including detection time, correct rate and so forth. In [5], authors adopt SVM ensembles for traffic incidents detection and obtain good performance. In [24], SVM are employed to confirm whether an incident occurred by means of vehicles' kinetic information analyses. Bayesian networks (BN) has been employed for incidents detection [6][7]. In [6], the authors proposed a BN-based module for relevant traffic knowledge storage and management, and then decide whether the traffic incidents occur in freeways. They also applied the BN to detect incidents in arterial road [7].

TSA approaches treat traffic data changing with time series in a regular pattern. An incident can be estimated when some changes inconsistent with the normal patterns happened. Cumulative sum (CUSUM) algorithm has been applied to process collected traffic data for AID in freeways [8], [9]. Standard normal deviation (SND) methods are also adopted in this problem to explore some situations involving obvious differences with normal patterns [8]. [10], [11] employ partial least squares regression (PLSR) and find that PLSR can have faster performance compared with SVM. Double exponential smoothing (DES) methods are also utilized for traffic variables estimation to confirm whether an incident occurs [25].

There are numerous other approaches applied in AID as well, with the most representative are decision trees algorithms [12], [26], patterns recognition techniques [13] and wavelet-based methods [27].

Hybrid approaches refer to a combination of techniques grouped into more than one categories mentioned above. ANN and PLSR have been integrated in this problem [14]. TSA and ML-based techniques has been mixed together to detect suspicious situations in freeways [15].

III.PROBLEM FORMULATION

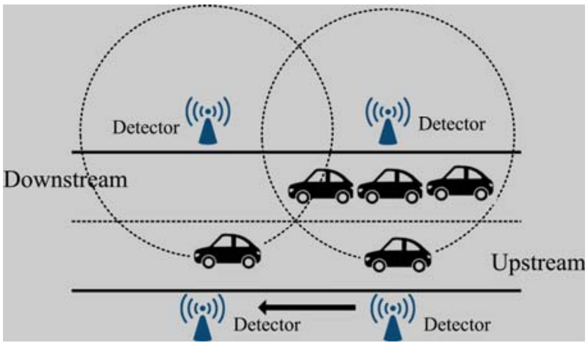

Fig.1 A Road Scenario

The problem of AID is how to detect the considerable abnormal traffic situation from plentiful and dynamic changing traffic states, with only two results, incident occurred or incident-free. It is similar to the binary classification problem. Our objective is to find the red line to separate these green circles and blue tri-angles into different sides. In this way, when some traffic variables are detected real-timely and inducing a traffic situation deviation with regular traffic patterns, we can utilize the red line to confirm which side these traffic variables should take place in.

To model this problem, we would consider a detector-equipped freeway road scenario (see Fig.1) which divided into several segments due to detector's detection range. We assume that when each vehicle comes into a road segment, the corresponding detector can sense its existence successfully.

We represent traffic variables in different segments as vectors $\mathbf{x(i)}$, i=1,2,3,…,N, defined as a road segment label. Each $\mathbf{x(i)}$ has its own final result, defined as $y(i)$, $y(i) \in \{-1,1\}$. Our objective is to find a function $\mathbf{F}$ in the following expression.

$$F: \mathbf{x} \rightarrow y \qquad (1)$$

IV.OUR AUTOMATIC INCIDENT DETECTION APPROACH

In this section, we propose our AID approach based on the above established model. The work flow of our approach has been depicted in Fig.2. There are four steps, data collection, data preprocess, data utilization in SVM model, and situation determination.

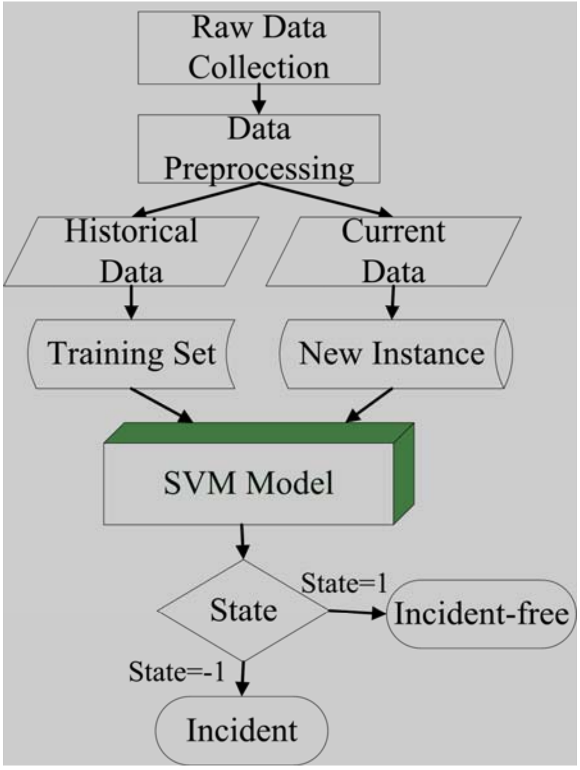

Fig.2 Work Flow of our SVM-based AID Approach

When detecting traffic data, sensors equipped roadside are usually the popular choices due to its convenient deployment and maintenance, such as wireless sensors. After the real-time traffic data collected, they need to be preprocessed in order to adapt to SVM model. In this model, when an incident happens in a segment, traffic volume of this segment and following segments would grow rapidly, with tangible reduction in the segments ahead. Similar change trends would occur on segment occupancy. In terms of average traffic flow speed, the speed of this segment and following segments would decrease obviously, with distinct improvement in the segments ahead. Hence, we decide to treat both traffic volume difference and speed difference between current segment and segment ahead as input variables for the SVM model, which means the data preprocess part should finish this job when receive all the traffic data collected. Moreover, we treat the historical data as training data, and the real-time detected data as a new instance. The detail mechanism of the SVM model would be presented in following. Based on the output of the SVM model, we can confirm whether an incident happens.

Based on the analysis mentioned above, vector **x(i)** has two elements, traffic volume difference between segment *i* and segment ahead *i+1*, defined as *tvd(i, i+1)*, and speed difference between segment *i* and segment ahead *i+1*, defined as *sd(i, i+1)*.

$$x(i) = (tvd(i, i+1), sd(i, i+1))^T \quad i = 1, 2, 3, \dots, N \tag{2}$$

The objective is find a maximum-margin hyper-plane $\omega \cdot \mathbf{x} + b = 0$, which divides the variables with *y(i)* equal to 1 from those with its value equal to -1. With a non-negative slack variable ξ_i to avoid mislabeled instances, the optimization problem can be presented in formula (3).

$$\begin{aligned} \min_{\omega,b,\xi} \quad & \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N} \xi_i \\ s.t. \quad & y_i(\omega \cdot \mathbf{x_i} + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, 3, \dots, N \end{aligned} \tag{3}$$

The SVM soft margin maximization problem can be transferred to its corresponding dual problem. And its purpose is to find the optimal α^*, and $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$.

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle - \sum_{i=1}^{N} \alpha_i \\ s.t. \quad & \sum_{i=1}^{N} \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, 2, 3, \dots, N \end{aligned} \tag{4}$$

After optimal solution obtained, and

$$\omega^* = \sum_{i=1}^{N} \alpha_i^* y_i \mathbf{x_i} \tag{5}$$

$$b^* = y_i - \sum_{i=1}^{N} y_i \alpha_i^* \langle \mathbf{x_i}, \mathbf{x_j} \rangle \tag{6}$$

Thus,

$$F(\mathbf{x}) = sign(\omega^* \cdot \mathbf{x} + b^*) \tag{7}$$

The well-known sequential minimal optimization (SMO) algorithm is employed for this problem with cost O($n^{2.3}$) in training and cost O(v) in testing, where n is defined as number of data instances and v is defined as number of support vectors [28].

V.EXPERIMENTS AND ANALYSIS

A. *Experiment Data Preparation and Evaluation Metrics*

The traffic dataset used for experiments is derived from the publicly available I-880 database from the Freeway Patrol Service Project in California, USA [29][30]. This dataset includes the traffic data we demand for, such as traffic volume and speed. And they also include abundant incidents events, almost 45 lane-blocking incidents [15].

The most common and widely acknowledge evaluation metrics for AID are detection rate (DR), false alarm rate (FAR), and mean time to detect (MTTD). DR is defined as the proportion of correctly found traffic incidents in all traffic incidents, presented in formula (8). FAR is defined as the proportion of false decisions in all incident-free cases, and presented in formula (9). MTTD is defined as the average value of each periods cost from the moment a traffic incident happens to the moment the traffic incident detected, and presented in formula (10), N is defined as the total incident number.

$$DR = \frac{\#\ of\ correctly\ detected\ incidents}{\#\ of\ all\ incidents} \tag{8}$$

$$FAR = \frac{\#\ of\ false\ decisions}{\#\ of\ all\ incident\text{-}free\ cases} \tag{9}$$

$$MTTD = \sum_{i=1}^{N} \frac{T_{detection}(i) - T_{incident}(i)}{N} \tag{10}$$

B. *Experimental Design and Analysis*

In automatic incidents detection problems, we would prefer higher DR, lower FAR, and shorter MTTD, which leads to a multi-purposes problem. The three goals are difficult to achieve optimal solution simultaneously. A higher DR may cause higher FAR and longer MTTD. Hence, we evaluate the performance separately, DR versus FAR and MTTD versus FAR respectively. Since our approach is based on SVM and select different features in the training stage, we adopt two representative related works [4],[18] as comparative approaches.

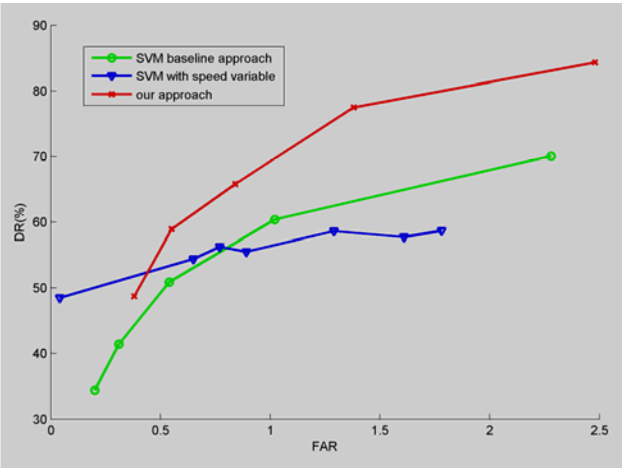

Fig.3 DR Comparison between the Three Approaches

Fig.3 presents the detection rate comparison between a SVM baseline approach [4], a SVM approach with speed variable [18] and our proposed approach. From the figure, we can observe that our approach can outperform the other two approaches in most cases. When the FAR lower than 0.5%, the SVM baseline approach presents the best performance. With the incidents number increases, all three methods have witnessed higher FAR, companied with higher DR. At that time, our approach obtains the best performance and expands the difference from the other two approaches.

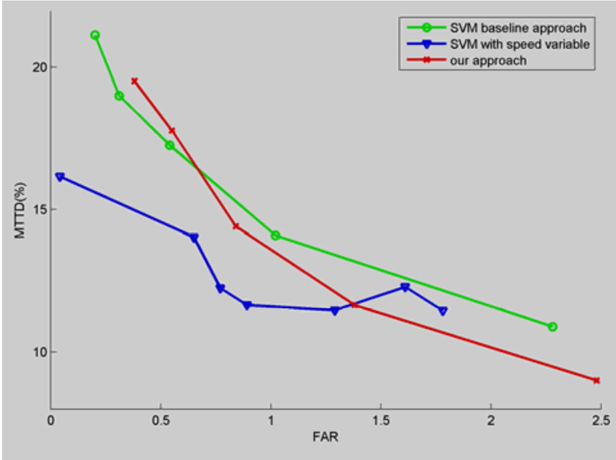

Fig.4. MTTD Comparison between the Three Approaches

Fig. 4 presents the mean time-to-detect comparison between the three approaches. From the figure, we can notice that with the three approaches have different performances when with different FAR. When the FAR is lower than 1.4%, the approach from [18] achieves best performance with much lower MTTD. When the FAR is higher 1.5%, our approach can outperform the other two approaches.

VI. CONCLUSIONS

Nowadays, traffic congestion problems have increasingly attracted people's attention due to the huge financial cost and air pollutions. Traffic incidents have been pointed out that plays an important part in traffic congestion. In this paper, we have presented an AID approach based on SVM with appropriate features, with traffic data detected by VANET techniques in a real-time manner. After several experiments conducted based on a real-world dataset, we confirm that our features selected can be beneficial for incidents detection, with higher detection rate and low mean time-to-detect with a certain level FAR, compared with two representative related works. In the future, we will optimize our work to further improve the detection rate, and we would pay efforts to optimize current approach in order to apply into urban areas.

ACKNOWLEDGMENT

This work is supported in part by Public Welfare Technology Applied Research Program of Zhejiang Province, China (2014C33108, 2015C33028, 2015C33074, 2015C33083), Research Fund of the Education Department of Zhejiang, China (Y201534553), Zhejiang Provincial Natural Science Foundation of China (LZ15F020001, LY13F010013, LY15F030004, LY14F020008).

REFERENCES

[1] J. Luk, C. Han, and D. Chin, Freeway Incident Detection–Technologies and Techniques, Sydney, Australia: Austroads, p. 72. 2010,

[2] X. Jin, R. Cheu, and D. Srinivasan, “Development and adaptation of constructive probabilistic neural network in freeway incident detection,” Transportation Research, 10C (2), pp. 121-147.2002,

[3] D. Srinivasan, V. Sharma, and K. A. Toh, “Reduced multivariate polynomial-based neural network for automated traffic incident detection,” Neural Netw., vol. 21, no. 2/3, pp. 484–492, Mar./Apr. 2008.

[4] F. Yuan and R. L. Cheu, “Incident detection using support vector machines,” Transportation Research Part C: Emerging Technologies, vol. 11, no. 3/4, pp. 309–328, Jun.–Aug. 2003.

[5] S. Chen,W.Wang, and H. VanZuylen, “Construct support vector machine ensemble to detect traffic incident,” Expert Systems with Applications, vol. 36, no. 8, pp. 10 976–10 986, Oct. 2009.

[6] K. Zhang and M. A. P. Taylor, “Towards universal freeway incident detection algorithms,” Transportation Research Part C: Emerging Technologies, vol. 14, no. 2, pp. 68– 80, Apr. 2006.

[7] K. Zhang and M. A. P. Taylor, “Effective arterial road incident detection: A Bayesian network based algorithm,” Transportation Research Part C: Emerging Technologies, vol. 14, no. 6, pp. 403–417, Dec. 2006.

[8] S. M. Tang and H. J. Gao, “Traffic-incident detection-algorithm based on nonparametric regression,” IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 1, pp. 38–42, Mar. 2005.

[9] H. Teng and Y. Qi, “Detection-delay-based freeway incident detection algorithms,” Transportation Research Part C: Emerging Technologies, vol. 11, no. 3/4, pp. 265– 287, Jun.–Aug. 2003.

[10] W. Wei, C. Shuyan, and Q. Gaofeng, “Comparison between partial least square and support vector machine for freeway incident detection,” 10th international IEEE conference on intelligent transportation systems, Doubletree Hotel, Seattle, Washington, USA, 2007.

[11] W. Wang, S. Chen, G. Qu, “Incident detection algorithm based on partial least squares regression,” Transportation Research Part C: Emerging Technologies, 16,pp. 54-70. 2008,

[12] H. J. Payne, E. D. Helfenbein, and H. C. Knobel, “Development and testing of incident detection algorithms,” Transportation Research Record, vol. 2, Report No. FHWA-RD-76-20, TRIS, pp. 316. 1976,

[13] J.F. Collins, C.M. Hopkins, and J.A. Martin, Automatic incident detection TRRL algorithms HIOCC and PATREG, TRRL Supplementary Report, No. 526, Crowthorne, Berkshire, U.K., 1979.

[14] L. Lu, S. Chen, W. Wang, and Z. Zuylen, “A hybrid model of partial least squares and neural network for traffic incident detection,” Expert Systems with Applications, pp. 4775-4784,2012

[15] J. Wang, X. Li, S. Liao, and Z. Hua, “A Hybrid Approach for Automatic Incident Detection,” IEEE Transactions on Intelligent Transportation Systems, vol.14, pp. 1176-1185. 2013,

[16] E.Gakis, D.Kehagias, D.Tzovaras, “Mining traffic data for road incidents detection,” 2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), pp.930-935, 2014.

[17] H. M. Wen, Z. S. Yang, G. Y. Jiang, and C. F. Shao, “A new algorithm of incident detection on freeways,” in Proc. IEEE Int. Conf. Veh. Electron.,pp. 197–202. 2001

[18] R. L. Cheu and S. G. Ritchie, “Automated detection of lane-blocking freeway incidents using artificial neural networks,” Transportation Research Part C: Emerging Technologies, vol. 3, no. 6, pp. 371–388, Dec. 1995.

[19] R. L. Cheu, and S. G. Ritchie, “Loop-Based Freeway Incident Detection Using Neural Networks,” Journal of the Institution of Engineers, Singapore, Vol. 35, No. 2, pp. 26-32. 1995,

[20] H. Dia and G. Rose, “Development and evaluation of neural network freeway incident detection models using field data,” Transportation Research Part C: Emerging Technologies, vol. 5, no. 5, pp. 313–331, Oct. 1997.

[21] S.G. Ritchie, and B. Abdulhai, “Enhancing the Universality and Transferability of Freeway Incident Detection Using a Bayesian-Based Neural Network,” Journal of Transportation Research part C, 7, 261-280, 1999.

[22] A. Karim, and H. Adeli, “Comparison of Fuzzy-Wavelet Radial Basis Function Neural Network Freeway Incident Detection Model with California Algorithm,” Journal of Transportation Engineering, ASCE, 128 (1), pp. 21-30. 2002

[23] S. I. Khan and S. G. Ritchie, “Statistical and neural classifiers to detect traffic operational problems on urban arterials,” Transportation Research Part C: Emerging Technologies, vol. 6, no. 5/6, pp. 291–314, Dec. 1998.

[24] Y. C. Ma, M. Chowdhury, A. Sadek, and M. Jeihani, “Real-time highway traffic condition assessment framework using Vehicle-Infrastructure Integration (VII) with Artificial Intelligence (AI),” IEEE Transactions on Intelligent Transportation Systems, vol. 10, no. 4, pp. 615–627, Dec. 2009.

[25] A. R. Cook and D. E. Cleveland, “Detection of freeway capacity-reducing incidents by traffic-streams measurements,” Transp. Res. Rec., J. Tranp. Res. Board, vol. 495, pp. 1–11, 1974.

[26] K.N. Balke, An evaluation of existing incident detection algorithms, Research Report, FHWA/TX-93/1232-20, Texas Transportation Institute, the Texas A&M University System, College Station, TX, November 1993.

[27] Y.-S. Jeong, M. Castro-Neto, M. K. Jeong, L. D. Han, “A waveletbased freeway incident detection algorithm with adapting threshold parameters,” Transportation Research Part C, 19, pp. 1-19. 2011

[28] J. Platt, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” Adv. Large Margin Classif., vol. 10, no. 3, pp. 61–74, Mar. 1999.

[29] K. F. Petty, Incidents on the freeway: Detection and management, Ph.D. dissertation, Dept. Elect. Eng. Com. Sci., Univ. Calif. Berkeley, Berkeley, CA, USA, 1997.

[30] H. Al-Deek, Y. Fawaz, H. Noeimi, K. Petty, D. Rydzewski, K. Sanwal, A. Skabardonis, and P. Varaiya. [Online]. Available: http://ipa.eecs.berkeley.edu/~pettyk/FSP/1995.