A Measurement-based System for TMC Performance Evaluation

Final Report

Craig Rindt and Will Recker

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600

December 20, 2010

STATE OF CALIFORNIA DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV. 10/98)

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA11-0975		

4. TITLE AND SUBTITLE	5. REPORT DATE
Transportation Management Center (TMC) Performance Measurement System	12/16/2010
	6. PERFORMING ORGANIZATION CODE None
7. AUTHOR(S) Will Recker, Craig Rindt	8. PERFORMING ORGANIZATION REPORT NO. None
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute of Transportation Studies University of California, Irvine Irvine, CA 92697-3600	10. WORK UNIT NUMBER 3763 11. CONTRACT OR GRANT NUMBER ***65A0252*** ***(Project ID#: 00-0000-0582)***
12. SPONSORING AGENCY AND ADDRESS California Department of Transportation Division of Research and Innovation, MS-83 1227 O Street; Sacramento CA 95814	13. TYPE OF REPORT AND PERIOD COVERED Final Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTAL NOTES

None

16. ABSTRACT

17. KEY WORDS	18. DISTRIBUTION STATEMENT	
	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. SECURITY CLASSIFICATION (of this report) None	20. NUMBER OF PAGES	21. PRICE Not Applicable

Reproduction of completed page authorized

Executive Summary

This document reports on the results of contract #65A0252, *A Measurement-based System for TMC Performance Evaluation*. This project developed a TMC performance evaluation system that allows TMC managers to evaluate the performance of various bundles of TMC technologies and operational policies by mapping their effects onto events in the system that can be measured using existing surveillance systems and daily activity logs. The work included tasks to develop a method for quantifying the relative benefits of TMC operations as well as tasks to provide a user friendly interface for performing benefits analysis of past incidents using available data sources. The system was deployed atop the CTMLabs web architecture and is available to approved users as part of the secure, authenticated, CTMLabs website.

[**(crindt)** Drop completed work, create summary and abstract.]

Completed Work

In Q2-2008, the team began a discovery process to determine resources and data available for implementing the performance evaluation system. We identified the major data available (freeway loop data, TMC activity logs, CHP CAD feed) as core sources and obtained representative samples for analysis. We also started a review of the performance evaluation methodology with the input of Caltrans TMC personnel.

In Q3-2008, we settled on the delay calculation methodology to be used as the primary analysis tool for the performance evaluation project. As part of this process, we developed a prototype software implementation using `perl` scripting that takes inputs from local copies of PeMS freeway data and the TMC activity log databases, generates a mathematical program to determine the impacted freeway region for given incidents, and solves that program using the `GAMS` solver. This software was used to successfully analyze a set of incidents identified by D12.

In Q4-2008, the team worked on finalizing the complete performance evaluation system prototype. This work included creating the ability to accept incident data from the CHP CAD XML feed, which is archived for cross referencing with the Caltrans activity log. Further work was carried out on encapsulating the core delay calculation algorithm so that it can be integrated into a user-driven performance evaluation application. Discussions during the quarter with the TMC team revealed that modifications can be made to the activity log in order to better support the performance evaluation system. The team engaged in a fact-finding exercise to better understand the current TMC architecture as well as the likely evolution of that system going forward. Based upon these assessments, the team developed a set of recommendations for upgrading the activity log.

Work during Q1-2009 was disrupted by a suspension of this contract due to California's budget crisis. Work carried out prior to the suspension of the contract focused on linking all information sources for the core performance evaluation engine developed in earlier work. A prototype web-based interface for viewing performance evaluation results was developed, but this interface has not yet been connected to the analytical back end due to the suspension of work on the contract. Additional work included developing interfaces for ITS's ATMS database, which will replace the static copy of sensor data that was used in initial development. Switching to the ATMS database will eventually speed deployment in the TMC.

Work during Q2-2009 continued on two fronts: developing and deploying a web-based interface for viewing performance evaluation results and the back-end system that will perform the performance evaluation. The prototype visually displays incident conditions and estimated impacted regions using a time-space block diagram of impacted facilities, similar to the figures generated in the original research. Work also continued on developing an efficient database containing rolling 1-year of historical sensor measurements from the ITS ATMS database. Work on linking the back-end data sources and the interface prototype was carried out off-line, but used data formats (JSON) that will be identical to those eventually produced by the back-end processing. Efforts to directly link the live data sources (ATMS, TMC activity log, and the CHP CAD XML feed) to the front end were put on hold until the revisions to the log discussed during November 2008 are completed by subcontractors during Q2-2009.

Work during Q3-2009 focused on three areas. On the modeling side, signficant progress was made in developing a more rigorous approach to converting sensor data into evidence of incident impacts. This work uses more advanced statistical methods to identify behavior at each sensor that differs from mean patterns. It is anticipated that this work will replace the existing threshold-based approach to generating evidence. Work also continued on readying the database to support the performance evaluation model and interface. These efforts are being synchronized with work on the related Safety project that relies on similar datasets. Additionally, progress was made on developing a consistent analytical representation of the D12 network to support the modeling of network effects. Toward this end, the freely available openstreetmap dataset was imported into the Testbed database for use in modeling and visualization. This will give the Testbed a map dataset that is unencumbered by license restrictions and which can be extended to meet Testbed needs in the future. Finally, work on the front-end interface continued, including the preliminary development of a map-based front end for viewing incident histories.

In parallel with the above efforts, the team continued to collaborate on the development of the activity log upgrade which will better support performance evaluation as it goes live. The team also engaged in a fact finding interview with TMC staff to document TMC processes used during incident management.

Ongoing and Planned Work

Going forward, work is slated to continue in several areas. We are exploring some better methods to use in the incident impact model (see section 3.1) for identifying when traffic conditions are statistically significant from the norm. Notably, we are working with members of the Department of Computer Science to apply Bayesian techniques to the analysis of traffic state data. We anticipate that these approaches will prove more robust than the course techniques currently being used.

In parallel to the above effort, we will begin developing an implementation of the TMC impact model (described in section 3.2).In this task, we will use a derived theoretical relationship between response time and delay magnitude and compare these predictions to the estimated total delays from the incident impact model for a subset of incidents managed by the D12 TMC. The goal is to demonstrate that the theoretical relationship generally holds and can therefore be used as a basis for predicting TMC benefits.

Finally, we will also bring the TMC interface to the performance evaluation system online. Initially, this will focus on the display of previously analyzed incidents (a prototype interface is shown in section 4.4.2).Fundamental to this effort is integration of the incident impact and TMC impact models into the Testbed's Enterprise Service Bus architecture that is currently under development. Furthermore, the system must be connected to the two main sources of data that support the analysis: the ATMS database (or UCI's copy of it) and the activity log database (or a shadowed version of it). Once these components are integrated, the web-based interface can be connected to live data and released for initial testing by potential users.

Contents

List of Figures

List of Tables

DISCLAIMER STATEMENT

Chapter 1

Introduction and Background

Caltrans has a significant investment in Transportation Management Center (TMC) throughout the state. These operations centers are tasked with maintaining the safety and efficiency of California's highways by actively managing disruptions to the system caused by anticipated and unanticipated events that impact the available capacity and/or the demand to use individual facilities. Presently, however, no comprehensive methods are available to quantify the benefits of existing TMC deployments. This research project developed a web-based TMC Performance Evaluation (TMCPE) application that addresses this problem. The system allows TMC managers to evaluate the performance of various bundles of TMC technologies and operational policies by mapping their effects onto events in the system that can be measured using existing surveillance systems and daily activity logs. The resulting tool provides managers with the long needed capabilities to:

- justify valuable technology, personnel allocations, and maintenance costs,
- identify technologies that aren't meeting their initial promise, and
- identify gaps in current operational strategies that might be filled with new technology deployments.

The evaluation method used considers delay savings that are attributable to specific TMC actions. All computations are based on direct measurement of the system using available sensors and do not rely on speculative simulation models requiring extensive assumptions.

The system is implemented using a modular software architecture that can interface with a variety of existing or planned systems used by Caltrans. The initial deployment focused on delivering a reporting system accessible from the recently deployed website for the California Traffic Management Laboratories (CTMLabs), formerly known as the California Advanced Transportation Management System (ATMS) Testbed. Possible future deployments could integrate the core performance evaluation models with real-time TMC management systems to assist operators with prioritizing response.

1.1 Performance measures of highway systems

Recent years have seen a significant push toward better evaluation of highway systems using data-driven performance measures. A number of recent federally-sponsored reports offer comprehensive overviews of current performance monitoring approaches and recommended practice for transportation management agencies. A common theme of developing an effective performance measurement system for operations is an early and complete definition of the goals of the traffic management system. These goals structure specific efforts in defining performance measurement tasks and applying the results to improve operations. Once system goals are defined, they must be translated into metrics that are obtainable from available (or feasible) data collection systems. Finally, the metrics should be integrated into management and operations to consistently evaluate the effectiveness of the system and, ideally, use that feedback across various time-scales to improve performance over time. In the next section, we discuss how to apply this general method to the analysis of TMC performance.

1.2 Benefits of the TMC

Traffic management centers perform a range of functions to meet broad transportation management goals. We use the taxonomy offered by Park (2005) to characterize TMC functions into three general areas: traffic monitoring, travel information, and event management.

1.2.1 Traffic monitoring

Traffic monitoring includes functions where transportation systems must be monitored in real-time, such as traffic flow monitoring and evaluating traffic conditions. To date, Caltrans has made significant progress in traffic monitoring. Data from the state's sensor network is available within and across districts—notably through the Performance Measurement System (PeMS) (Choe et al., 2002), which offers system-wide measures of effectiveness on traffic flow as well as the health of the sensor system itself.

1.2.2 Travel information

Travel information programs refer to programs available to the public that provide information on current traveling conditions. In California, a number of systems disseminate travel information to the traveling public through various public channels (traditional media, 511 systems, changeable message signs, traffic web sites, etc.) so that they can make more informed travel decisions.

Metrics on the benefits of this information dissemination are generally not available except in vary broad terms from the results of travel behavior surveys in response to information. Understanding more clearly how travelers respond to specific information provided by the TMC could identify critical contributions of TMC operations that heretofore are undervalued.

1.2.3 Event management

Events managed by the TMC may be (1) random (e.g., incidents), (2) recurring (e.g., congestion from prevailing traffic patterns) and (3) scheduled (maintenance or special events). Generally, such events can impact the normal operating capacity of the system or can produce demands that are significantly distorted from the recurrent patterns for which system capacity and control has been optimized. Event management—particularly of non-recurrent congestion caused by incidents—is likely the most tangible benefit offered by the TMC. Procedures vary across districts and are tuned to the specific capabilities of each, but urban districts with TMCs have specialists who actively manage disruptions to the system. Nonetheless, event management is an area of continued development for Caltrans.

The typical metric for evaluating TMC event management processes is delay savings. A number of methods have been proposed to estimate total non-recurrent congestion delay. They can be classified into three groups:

1. analytical methods using deterministic queuing diagrams (Goolsby and Smith, 1971; Chow, 1974; Morales, 1987; TRB, 1994; Lawson et al., 1997; Erera et al., 1998; Skabardonis and Geroliminis, 2004, e.g.,)

2. kinematic wave (i.e., shockwave) theory (Chow, 1974; Wirasinghe, 1978; Heydecker, 1994; Al-Deek et al., 1995, e.g.,)

3. heuristic methods (Skabardonis et al., 1996, 2003, e.g.,)

The results from these approaches have been applied both alone and in concert with each other. But, since virtually none of these studies uses a source of incident data with descriptive variables and reporting techniques in common with the others, comparisons between different methods are difficult (Goolsby and Smith, 1971), and their spatial transferability is limited. To date there has been only limited research into models that can estimate how long, how far, and to what extent any particular incident will affect traffic and, by extension, little research into understanding how TMCs perform at minimizing the impact of such incidents. The focus in this project has been the development of incident-specific delay-estimation metrics that can be used to evaluate TMC performance.

1.3 Data sources

Performance evaluation requires data. Specifically, data is needed about what is happening in the freeway system and data is needed about actions taken by the TMC and other agencies. In its role as an information broker for Caltrans, the TMC has access to a range of information that can meet these needs in whole or part. We summarize them in the below as they dictate the types of performance evaluation analyses that are possible

1.3.1 System data

The real-time traffic data collected by Caltrans loop detectors stations are the primary source of information regarding the actual performance of the freeway system during incidents. A number of sources supply this data in a manner that can be used for this project. It should be noted, however, that a local archive of the data must be maintained as the delay calculations rely on customized queries for relevant sensor data. These queries are not generally available from the systems listed below.

1.3.1.1 PeMS

The Caltrans Performance Measurement System has become the *de facto* standard for processed freeway sensor data in the state of California. The original work that served as the basis for the core of the TMCPE system developed in this research (Recker et al., 2005) used the PeMS database to obtain a year of 5-minute data from every sensor in Orange County. This data was archived in a local database to support the prototyping of efficient database queries necessary for the analysis. PeMS offers access to this data archive that is scriptable and therefore could be incorporated as the primary data source for archived performance in the system for the current effort.

1.3.1.2 The CTMLabs sensor archive

The CTMLabs provide an ideal implementation platform for accessing detector data because the CTMLabs receive identical data to that available in the TMC by way of the CTMLabs real-time data intertie and the associated database. As such, development can proceed in the CTMLabs and later be transferred seamlessly to operation in Caltrans, District 12 (D12).

1.3.1.3 Caltrans Distrct 12 XML/SOAP interface

For greater transferability, the TMCPE system can also make use of the *devices* implementation in the District 12's SOAP interface. This interface implements portions of the center-to-center interface of the National ITS Architecture standard—including the Vehicle Detector Stations (VDS) interface standard which streams 30-second sensor readings from all detectors in the district. In order to make use of this interface, the TMCPE system would need to build its own archive of sensor data received. Since this is already being done as part of the CTMLabs sensor archive, reproducing this effort to support the TMCPE application is not necessary.

1.3.2 TMC actions

The actions of the TMC and other event response personnel are recorded by available from several sources as outlined below.

1.3.2.1 CHP iCAD log (XML feed)

The California Highway Patrol (CHP) Intelligent Computer Aided Dispatch (iCAD) system logs all roadway incident reports including: incident date, CHP iCAD identification number (ID), CHP iCAD time stamp, incident type, and incident location. This data is not presented in a canonical format and therefore is not necessarily easy to parse automatically. Nonetheless, it is the only source for raw information regarding multi-agency incident response.

The functions provided by the TMCPE project require a local copy of the key data from the CHP iCAD log—particularly the geographic location of incidents. If the geographic data is added to the Activity Log (see section 1.3.2.3 below), the direct dependence of the TMCPE system on the CHP iCAD log can be removed.

1.3.2.2 D12 SOAP interface (events, devices)

The D12 SOAP interface also streams event data from the TMC and control states of Changeable Message Signs (CMSs) and Ramp Metering Stations (RMSs). The event data could be tapped to obtain incident verification details, while the CMS and RMS settings could be parsed to infer control actions carried out by operators. Ideally, however, any control actions would be tied to specific TMC processes and their associated goals (see section 6 for further discussion on relating actions to TMC processes).

1.3.2.3 TMC activity log

The D12 TMC activity log records a range of actions carried out by TMC operators that they manually record using a web-based interface to the log database. Transactions in the database are cross referenced with CHP iCAD IDs, which allows additional data from the iCAD system—such as the incident's geographic coordinates—to be associated with each incident under analysis. At the beginning of this project, only a limited number of database entries were reliable enough to be parsed automatically and thus be used directly by the TMCPE system. These were the `OPEN/CLOSE INCIDENT` and `SIGALERT BEGIN/END` entries. Additional entries relating to the status of changeable message signs and construction closures were useful for cross checking, but generally were not able to be used in the calculations described in chapter 3 in their present form. Nonetheless, the activity log's comprehensive nature and integration into TMC processes in D12 made it the natural choice for obtaining information about incident management in the TMC.

1.3.2.4 TMCAL

During the course of this project, Caltrans began work on developing the Traffic Management Center Activity Logging (TMCAL) system to provide a common, statewide standard for recording information about managed incidents in the state of California. The users and developers of the D12 activity have participated in the specification of TMCAL to ensure that the functionality of the D12 TMC activity log is replicated in TMCAL. As such, we proceeded to develop of the TMCPE system using the D12 TMC activity log as the reference data source under the assumption that a later shift to the TMCAL system would be straightforward given their common ancestry.

1.4 Organization of this report

The remainder of this report is structured as follows. Chapter 2 describes the processes used in the Caltrans D12 TMC during the management of accidents, which is the initial target of this performance evaluation effort. Chapter 3 details the methods being used for performance measurement—with a focus on the calculation of delay attributable to specific incidents as well as a discussion of how to model the impacts of TMC actions on the dynamics of such incidents. Chapter 4 describes the system architecture that was developed to deploy the system as an operational tool for the TMC as part of the CTMLabs. In chapter 5, we discuss representative results from the system and we conclude in chapter 6 with some recommendations for continued development of the TMCPE system.

Chapter 2

TMC Activity

At the onset of this project, the district 12 TMC employed locally-developed TMC activity log software (section 1.3.2.3) to record traffic management activity carried out by the TMC. This software was developed over a number of years with a variety of purposes in mind. Though the initial analyses have been conducted using this existing logging data, the data is unsuitable for general purpose automated processing to identify critical events in TMC operations that impact system performance. As such, work was initiated under this contract to improve the activity logging system to better meet such performance monitoring needs going forward.

A critical first step in this process was the identification of core events in incident management that can be directly recorded by the activity logging system through a set of interviews and communications with district 12 staff and contractors. The following summarizes the TMC processes used to manage accidents occurring in the managed system. Note that this effort was carried out in parallel to a number of efforts in district 12 and beyond to capture and organize event management, including D12's event management system development and the Caltrans-wide effort to develop the TMCAL system. Any findings below should be considered one characterization of the larger problem of organizing TMC processes to best manage, monitor, and evaluate Caltrans traffic management in general, and TMC activity in particular.

Based upon our interviews, incident management by the TMC can be described as four major subprocesses that are traversed sequentially as shown in figure 2.1 Each subprocess dictates a series of actions to be carried out by the TMC. The transition between the subprocesses is governed by a series of events whose dynamics are dictated by events in the field.

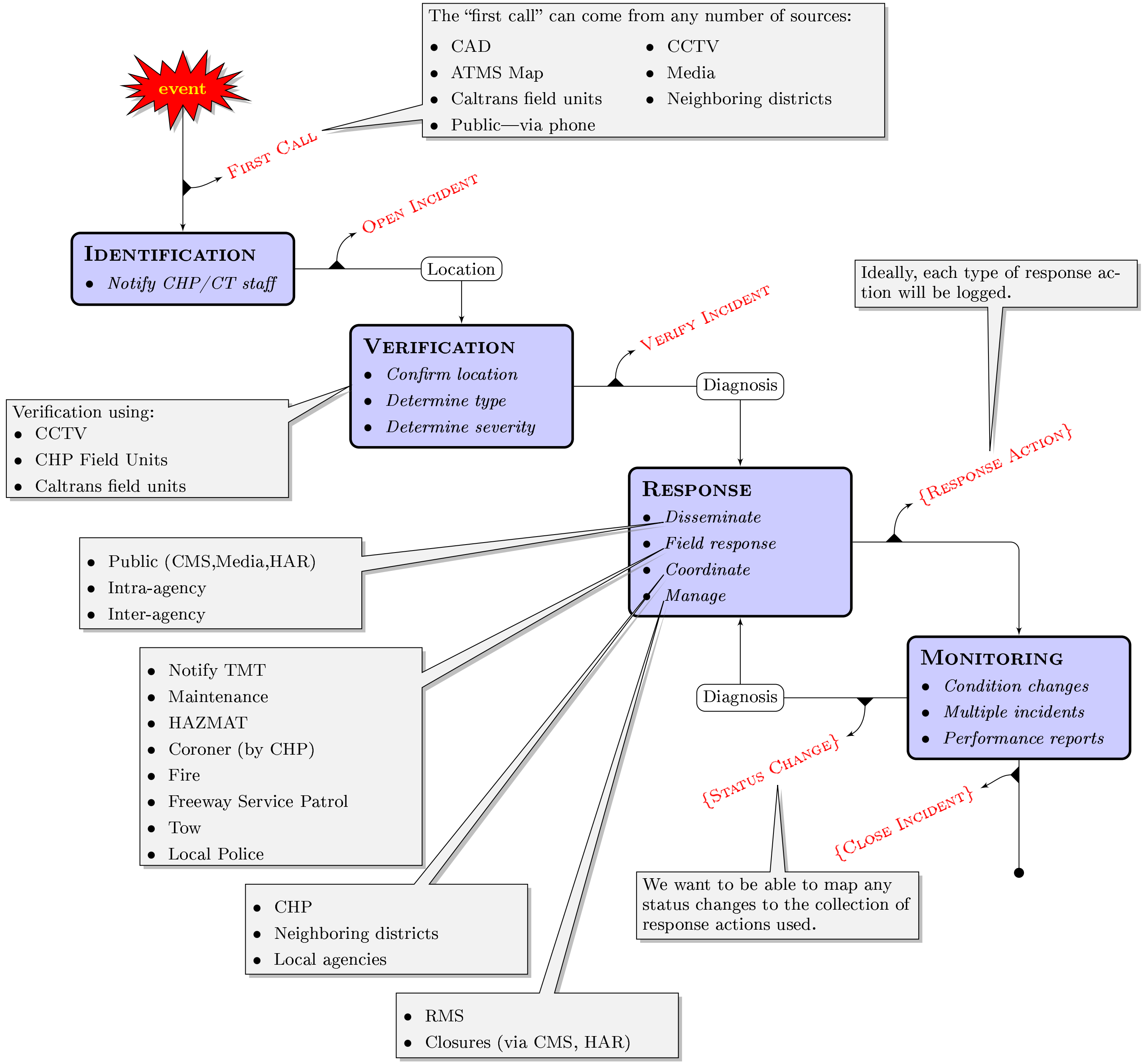

Figure 2.1: TMC process flowchart for accident management. Four major subprocesses dictate a series of actions to be carried out by the TMC. The trajectory through the subprocesses is governed by a series of events whose dynamics are dictated by events in the field.

2.1 Identification

TMC management of an incident always begins with the First Call to the TMC, which notifies operators that there has been a disruption to what is considered “normal” operating conditions. From an operational perspective, these disruptions can affect the capacity of particular sections of roadway (as in a traffic accident), the prevailing demand to use particular stretches of roadway (as in a special event), or some combination of the two. Our emphasis for the current discussion is capacity-reducing accidents.

The First Call can come from any number of sources including:

- CHP iCAD
- ATMS software map
- Caltrans field units
- Public—via phone
- CCTV
- Media
- Neighboring districts

The vast majority come by way of the CHP iCAD system and the local CHP liaison to the TMC.

When a First Call is received, initial details are collected about the location, type, and likely impact on the system. TMC staff is notified and a decision is made to manage the incident, which we term an Open Incident event, at which point the process transitions to the Verification stage.

2.2 Verification

The purpose of verification is to identify and confirm the details of the incident to support response including the location, type, severity, and any other special conditions that are necessary to support TMC operations. Strictly speaking, the TMC cannot respond to the incident until its characteristics are known. This is particularly true in situations requiring special response, such as incidents requiring hazardous material spills. Since total queuing delay is proportional to the fourth power of the response time (see section 3.2.1), the speed with which verification is carried out can be a significant determinant to the quality of the response and the subsequent degree of disruption experienced by users of the system.

The TMC has a number of resources to support verification including:

- CCTV (both those owned by Caltrans and those owned by partner agencies)
- CHP field units
- Caltrans field units

TMC operators use information collected via resources to characterize the incident—which we term a diagnosis—for the purpose of formulating a response. Once a diagnosis is produced, verification is complete and the TMC moves to the RESPONSE subprocess.

2.3 Response

The response subprocess is understandably the most involved activity carried out by the TMC. The collection of activities carried fall into four categories: Information dissemination, Field Response, Coordination, and Control.

2.3.1 Information dissemination

The sharing of information regarding the incident is one of the most critical actions carried out by the TMC. Information must flow to a number of audiences. Among the most important recipients are the operators and management of other agencies—including neighboring Caltrans districts and management, whose actions may depend on or involve the affected facilities. Dedicated procedures exist for providing this communication, and for specifying when such communication must occur. For instance, headquarters is always notified when a SIGALERT is issued, updated, and terminated.

Similarly, the users of the system must be notified of disruptions. This serves the purpose of providing both higher quality of service to users, since lack of information degrades the user experience, as well as acting as a control mechanism to initiate unmanaged diversion (see below for prescribed diversion). A number of communications channels exist for notifying the public. The media is informed through various information feeds that are provided directly or indirectly by Caltrans and its partner agencies. For instance, members of the media can access the public CHP iCAD feed as well as speed information derived from Caltrans freeway sensors and exposed to various traffic information providers by via the ATMS software system. Issuance of a SIGALERT (by the CHP) or a Traffic Advisory (by

Caltrans) triggers specific interactions with the media. Information can be provided directly to users in the vicinity of the incident using the fixed or mobile CMS and Highway Advisory Radio (HAR) systems.

2.3.2 Field response

The field response to a particular incident is governed by the diagnosis generated during the VERIFICATION process. The Caltrans TMC is responsible for managing the actions of Traffic Management Teams (TMTs), HAZardoes MATerial (HAZMAT) crews, maintenance crews, and the coroner (managed by the CHP) that are dispatched to mitigate the impacts of an incident on traffic operations and system infrastructure. These actions are critical to restoring the performance of the system to normal and are therefore time critical events that must be logged effectively to support performance evaluation.

2.3.3 Coordinate

Though the transportation system is divided into multiple jurisdictions for management purposes, its global performance is linked by the actions of other agencies, most notably the CHP, neighboring Caltrans districts, and neighboring municipalities. From interviews with TMC staff, coordination with CHP is fluid, with the CHP and the TMC working in concert to manage incidents. Coordination with neighboring districts also involves fluid communications regarding actions and responsibilities. Coordination with municipal entities is limited except in the case of severe incidents. In those instances, response coordination is typically governed by pre-defined coordinated response plans that govern actions, for instance, in the case of a freeway closure and subsequent diversion to a municipality's surface street network.

2.3.4 Control

Active traffic control actions taken by Caltrans are typically limited to the most severe incidents. In those cases, Caltrans may institute a freeway closure and prescribed diversion. In very rare instances involving long-term closures, settings of field traffic controllers (such as time-of-day ramp metering settings) may be modified to account for shifts in prevailing traffic flow.

2.4 Monitoring

Each discrete response activity described above has an explicit or implicit intended outcome that would, in an ideal system, be directly measurable via the monitoring subprocess. In reality, the impact of only a subset of the response actions can be measured directly. For instance, the efforts of CHP, TMTs, etc. to clear the roadway can be recorded directly in the TMC activity logs as lanes restored. The impact of information dissemination, on the other hand, is very difficult to infer from the course sensor data available. The monitoring process should also track when additional event(s) occur that may impact the mitigation efforts for a prevailing incident. Management of multiple incidents may be handled separately or combined into a single management problem.

Generally, the monitoring process should keep track of changes to the system as they relate to the incident in question. Changes might include arrival of TMTs or other response crews, lane clearance, completion of HAZMAT efforts, length of queuing (for validation purposes) and any other information that is potentially relevant to incident post analysis. Each such (verified) STATUS CHANGE could produce an updated diagnosis and possibly changes to the TMC response, as indicated by the feedback loop in figure 2.1.

2.5 Practical Concerns

The TMC is a complex environment with a vast array of hardware and software collecting, managing, and disseminating information regarding the performance and operation of the Caltrans highway network. It is notable that a number of these systems have functions that occasionally overlap. For instance, the activity logging software is the *de facto* interface for TMC operators in D12 engaged in traffic management. However, a subset of TMC actions, like setting the CMS messages are carried out via the ATMS software system. Logging of CMS messages in the activity log must (currently) be done manually. Furthermore, Caltrans continues to actively develop an expert system for event management to improve event management processes by formalizing them and mandating that certain information is recorded and that particular actions are carried out.

In this complex environment, additional logging burdens are not likely to be well received. Ideally, many TMC actions would be automatically logged using messages passed between the various software systems (including the TMC activity log interface) to a logging service that maintains the backend logging database. With this in mind, we recommend a simple, extensible event tagging system that captures key points of the incident management process consistent with the above description. These event tags correspond to the red events in figure 2.1 and are summarized in table 2.1.

Event tag	Description
First Call	Logged when a TMC operator received the first report of an incident. Typically will be the first entry in the activity log with an associated CHP iCAD ID and should note the source of the notification.
Open Incident	Logged when the decision is made that a particular incident must be tracked and managed by the TMC.
Verify Incident	Logged when the details of the incident have been confirmed. Should include all sources (and technology) used to verify the characteristics of the incident.
Response Action:<Subtype>	Logged when a particular action is taken to mitigate the impact of an incident. The range of actions available for logging will vary with the situation.
Status Change:<Subtype>	Logged when a notable change to the managed event has occurred. For instance, arrival of TMT, restoration of a lane, etc. Informational messages for model validation are useful here too, such as noting the location of the back of the queue.
Close Incident	Logged when all management functions have terminated and traffic flow has returned to normal conditions.

Table 2.1: Core TMC events to be recorded by the activity log to support TMC performance evaluation

The Response Action and Status Change tags will ideally include a complete set of possible response/condition pairs that characterizes TMC actions and their effects. To the extent that the TMC Logging system can standardize logging to produce unambiguous tags for these events, the performance evaluation system will be able to reconstruct historical incidents for post-analysis. The next chapter describes that modeling methodology used in the TMCPE system.

[**(crindt)** Show better before and after logs]

Chapter 3

Modeling Methodology

To determine the relative impact of TMC technologies and processes on the performance of the system, we decompose the analysis into two general steps. In the first step, an *incident impact model* (section 3.1) estimates the degree to which an incident disrupts freeway operations in time and space (the "control volume") as well as the total delay in the control volume using sensor observations relative to "average" conditions. This delay estimate serves as a baseline condition for the analysis. In the second phase of the analysis, we use the *TMC impact model* (section 3.2) to explore how system performance would change in the absence of specific TMC actions by coupling observations and estimated changes in critical incident parameters with a model based upon queuing theory to estimate delays for the same incident if the TMC was absent.

3.1 Incident impact model

3.1.1 Section definition

A freeway section in this study corresponds to a mainline loop detector station. Section boundaries are defined by the middle of two detector stations as shown in figure 3.1. We assume that estimated speed at the station is representative of the speed for the corresponding section. Based on the sections and their corresponding detector stations, estimated speeds and densities for each section i are calculated for each 5-minute interval during the one-year analysis period—sections downstream of section i are labeled consecutively as $(i+1), (i+2), \ldots$ while those upstream are labeled consecutively as $(i-1), (i-2), \ldots$.

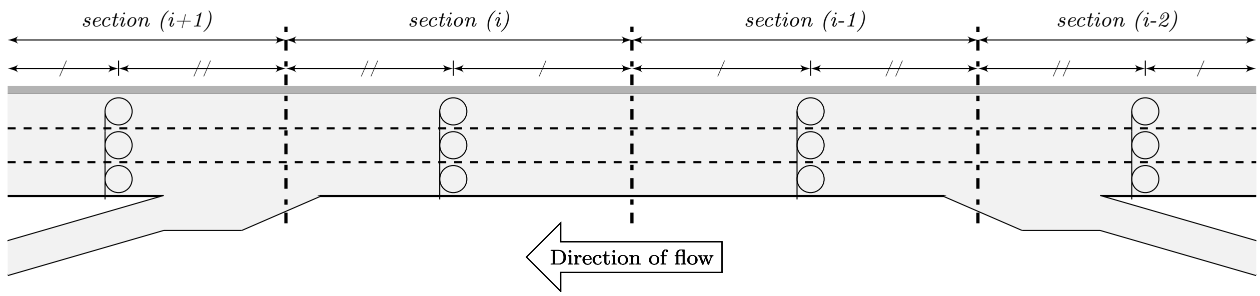

Figure 3.1: Section definition and corresponding detector location. Sections boundaries are formed by bisecting freeway segments between each pair of detectors. By convention, section numbers increase as you move downstream.

3.1.2 Stream state estimation using single loops

This method relies upon access to the prior one-year of historical traffic data— specifically: traffic counts and occupancy for each lane at each detector station every 30 seconds. Lane-by-lane traffic data is aggregated to 5-minute intervals at each detector station in order to obtain stable traffic data from each point.

Loop detectors in Orange County predominantly are single loop detectors that provide only traffic counts and occupancies; travel speeds need to be estimated from these measures assuming a so-called "g-factor," the summation of the average vehicle length and effective detection length (Hall and Persaud, 1989; Jia et al., 2001; Wang and Nihan, 2003). The average speed, $\overline{s}$, can be calculated as:

$$\overline{s}(^{\mathrm{mi}}/_{\mathrm{h}}) = g\left(\frac{q}{5280(^{\mathrm{ft}}/_{\mathrm{mi}}) \times occ}\right) \tag{3.1}$$

where

q = flow rate($^{\mathrm{veh}}/_{\mathrm{h}}$)

occ = vehicle occupancy(%)

g = g-factor

Applying appropriate g-factors is a key in estimating speeds from single loop detector data. There are various ways to estimate g-factors. The simplest approach is to use a single g-factor for all loop detectors in the study area. Although this is very common in practice, the approximation from this approach is too coarse for the objective of this study. An alternative is to use a single g-factor for each lane through the entire freeway system. This approach accounts for vehicle length varying by lane, but assumes the same proportion of heavy vehicles on every section of the freeway. Another approach is to find a single g-factor for every loop in each lane. This can be estimated by using flow and occupancy from some period for which speed can be assumed to be known; for example, flow and occupancy during 1:00AM-1:05AM, where the average speed can be assumed as a free flow speed, can be used to calculate the g-factor. This approach, however, does not account for the variation of the vehicle composition over time. Jia et al. (2001) noted that the g-factors for different loop detectors in the same district differ by as much as 100 percent, and the g-factors for the same loop can vary up to 50 percent over 24-hour period.

In this research, we account for spatial and temporal variations of the g-factors by calculating a g-factor representing each hour interval over the 52 weeks for each loop detector station as:

$$\hat{g}_i(t+1) = p \cdot \hat{g}_i(t) + (1-p) \cdot g_i(t+1) \tag{3.2}$$

$\hat{g}_i(t)$ = g-factor for section i for time step t

$g_i(t+1)$ = initial average g-factor for section i for time step $t+1$, obtained by assuming free-flow speed (75-mph) when the occupancy is lower than 0.06

p = smoothing parameter (0.9 in this research)

By applying the above procedure, representative g-factors for each hour of day are calculated for all loop detector stations. These g-factors provide basic information for speed calculation in this study. Based on these g-factors, speeds are calculated every 5 minutes for all 52 weeks.

3.1.3 Nominal speed distribution

For each freeway section j , for each specific 5-minute time interval t_m on each day of the week, a nominal total of 52 observations1 are available during the one-year observation period on which to base calculations of flow $q_j(t_m)$ and occupancy $\mathrm{occ}_j(t_m)$ under conditions that are incident free. For example, section j on Monday from t_m = 08:10 to 08:15 for 52 weeks is composed of 52 samples. Thus the speed associated with the n^{th} observation for any particular day-of-week/time interval/section combination can be estimated as:

$$\overline{s}_{jn}(t_m) = g_{jn}(t_m)\left(\frac{q_{jn}(t_m)}{5280 \times occ_{jn}(t_m)}\right); n = 1, 2, \ldots, K, \ldots 52 \tag{3.3}$$

where the subscript n is used to designate the estimate of the corresponding value based on the n^{th} observation. From these (nominally) 52 observations, a distribution of $s_{jn}(t_m)$ can be constructed, such as that shown in figure 3.6. Then, let $\Omega_{jm} = \Omega\left(\overline{s}_j(t_m), \sigma_{s_j}(t_m)\right)$ denote the

Time	Freeway Section i	$i-1$	$i-2$	$\cdots$	2	1
t_1	$\Omega_{i,1} = \Omega(\bar{s}_i(t_1), \sigma_{s_i(t_1)})$	$\Omega_{i-1,1} = \Omega(\bar{s}_{i-1}(t_1), \sigma_{s_{i-1}(t_1)})$	$\cdots$	$\cdots$	$\cdots$	$\Omega_{1,1} = \Omega(\bar{s}_1(t_1), \sigma_{s_1(t_1)})$
t_2	$\Omega_{i,2} = \Omega(\bar{s}_i(t_2), \sigma_{s_i(t_2)})$	$\Omega_{i-1,2} = \Omega(\bar{s}_{i-1}(t_2), \sigma_{s_{i-1}(t_2)})$	$\cdots$	$\cdots$	$\cdots$	$\Omega_{1,2} = \Omega(\bar{s}_1(t_2), \sigma_{s_1(t_2)})$
t_3	$\Omega_{i,3} = \Omega(\bar{s}_i(t_3), \sigma_{s_i(t_3)})$	$\Omega_{i-1,3} = \Omega(\bar{s}_{i-1}(t_3), \sigma_{s_{i-1}(t_3)})$	$\cdots$	$\cdots$	$\cdots$	$\Omega_{1,3} = \Omega(\bar{s}_1(t_3), \sigma_{s_1(t_3)})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
t_M	$\Omega_{i,M} = \Omega(\bar{s}_i(t_M), \sigma_{s_i(t_M)})$	$\Omega_{i-1,M} = \Omega(\bar{s}_{i-1}(t_M), \sigma_{s_{i-1}(t_M)})$	$\cdots$	$\cdots$	$\cdots$	$\Omega_{1,M} = \Omega(\bar{s}_1(t_M), \sigma_{s_1(t_M)})$

Figure 3.2: Base case of distributional properties for non-incident speeds $s_{jn}(t_m)$

set of parameters defining the distribution of speeds $s_{jn}(t_m)$ corresponding to this "incident-free" base case. (Here, we make no attempt to categorize these distributions among the set of well-known distributions, but rather assume, for simplicity, that they can be adequately defined by their means and standard deviations.) Then, associated with an incident known to have occurred on freeway section i during some time interval t_1 , we can compose a matrix of the parameters defining the distribution of base case conditions (i.e., conditions in which there is no incident) that can be expected to prevail for $t_m \geq t_1, m = 1, 2, \ldots, M$ (i.e., time intervals after the incident that occurs in section i at t1) for all upstream sections that could possibly have been affected by the incident, as in figure 3.2.

Similarly, associated with a known incident that occurs on freeway section i at time $t = t1$, we can construct a matrix of speeds $\hat{s}_j(t_m)$ based on observed loop data under the incident conditions, $\hat{q}_j(t_m), \hat{k}_j(t_m); j = i, i-1, i-2, \ldots, 1; m = 1, 2, \ldots, M$ as:

$$\hat{s}_j(t_m) = \frac{\hat{q}_j(t_m)}{\hat{k}_j(t_m)};\ j = i, i-1, i-2, \ldots, 1; m = 1, 2, \ldots, M \tag{3.4}$$

from which we can then compose a matrix of incident case speed conditions (i.e., conditions that were observed to prevail following the incident) as in figure 3.2. Using figure 3.2, we can identify the negative effects (i.e., speed reduction) of the incident schematically as shown in figure 3.4. The negative effect by the incident will be propagated from the incident section to upstream sections. Such a distinct discontinuity between non-congested and congested flow is known as a shock wave (Lighthill and Whitham, 1955). If the dot-shaded (yellow) area affected by the shock wave in figure 3.4 is identified, then the temporal and spatial impacts of the incident will be also determined. We approach this identification as a topology problem in which the objective is to construct the most probable set of cells that define the incident-affected temporal/spatial region, subject to constraints imposed by allowable shapes of the affected region. The following section describes the method for discriminating the regions

Time	Freeway Section i	$i-1$	$i-2$	$\cdots$	2	1
t_1	$\hat{s}_i(t_1)$	$\hat{s}_{i-1}(t_1)$	$\cdots$	$\cdots$	$\cdots$	$\hat{s}_1(t_1)$
t_2	$\hat{s}_i(t_2)$	$\hat{s}_{i-1}(t_2)$	$\cdots$	$\cdots$	$\cdots$	$\hat{s}_1(t_2)$
t_3	$\hat{s}_i(t_3)$	$\hat{s}_{i-1}(t_3)$	$\cdots$	$\cdots$	$\cdots$	$\hat{s}_1(t_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
t_M	$\hat{s}_i(t_M)$	$\hat{s}_{i-1}(t_M)$	$\cdots$	$\cdots$	$\cdots$	$\hat{s}_1(t_M)$

Figure 3.3: Observed incident speeds $s_{jn}(t_m)$

between non-congested and congested area due to traffic incidents.

3.1.4 Quantifying the total delay caused by an incident

3.1.4.1 Determining maximum extent of incident influence

Since data on neither the severity (such as the number of closed lanes by an incident) nor when the incident was cleared are directly obtainable from loop data, we first estimate the maximum possible extent of the shock wave by assuming the worst possible conditions—total blockage for some pre-specified time period. Thus, for any given incident occurring at section i at time t_1 , we compute the maximum number of upstream sections that could be affected by the assumed persistent total blockage at section i at time t_1 . Using this sort of data, we can schematically construct the "maximum area of interest" for any incident occurring at section i at time t_1 as the shaded (blue) area in figure 3.4. Based on this interpretation, the only data relevant to the current example of an incident occurring at section i at time t_1 can be restricted to cells in the shaded (blue) area. That is, the region of interest can be depicted as shown in figure 3.5.

3.1.4.2 Determining the congested region

Figure 3.5 shows two shaded areas: a dot-shaded (yellow) area and plain-shaded (blue) area. The cells in the dot-shaded (yellow) area represent speeds, $\hat{s}_j(t_m)$, that have been determined to have been affected by the incident; i.e., they have been deemed to be significantly lower than would be expected based on the distribution of non-incident base speeds for that particular section-time interval combination. The plain-shaded (blue) cells represent particular section-time interval combinations for which the observed speeds are not significantly different from non-incident conditions. That is, they are speeds that are determined from observation not to be affected by the pertinent incident, but within the maximum possi-

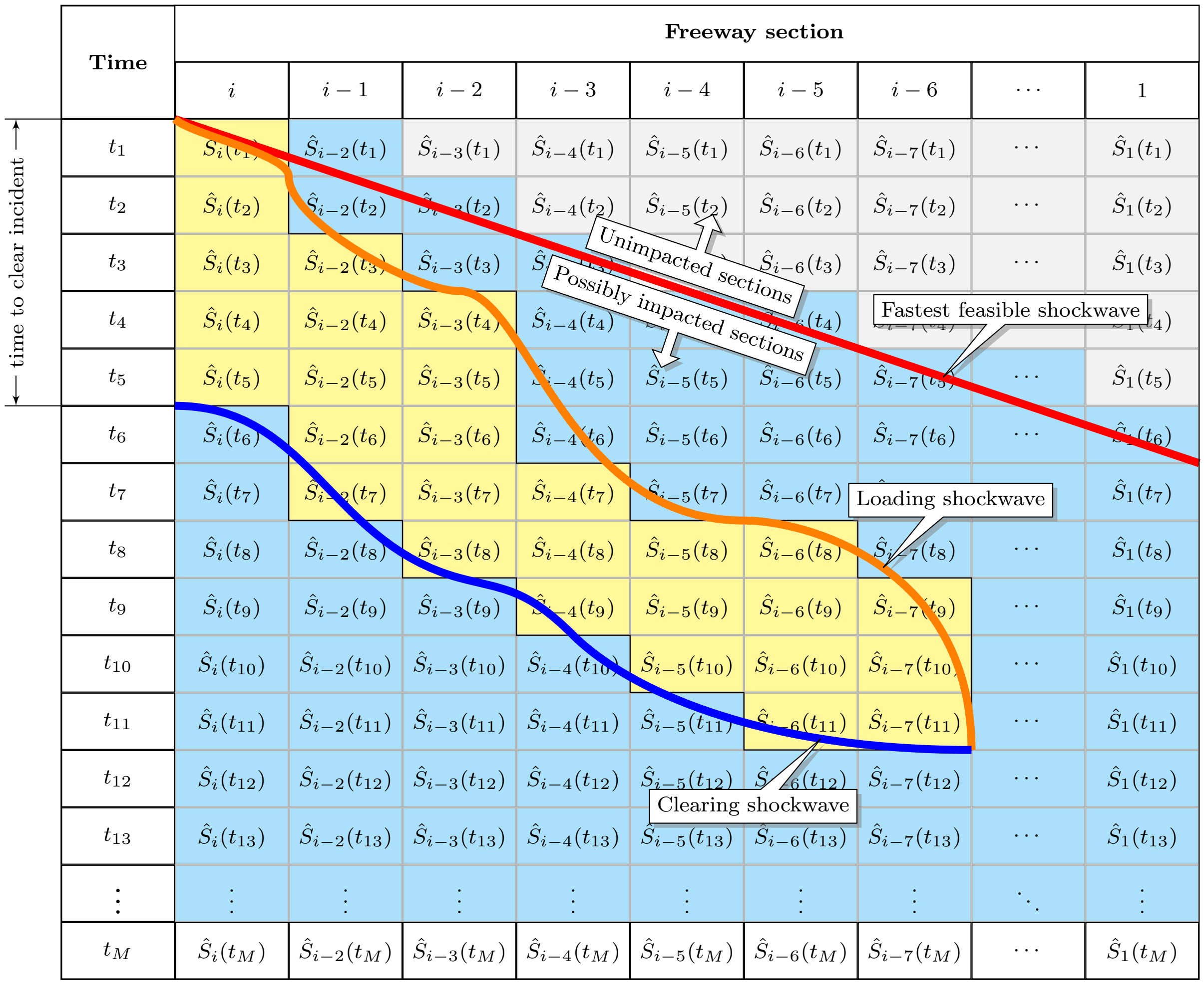

Figure 3.4: Incident impact projected into discrete space. The state of traffic flow is described in terms of discrete spatial (i) and temporal (t) sections as shown. An incident occurring at the downstream end of section i at time t_1 disrupts flow for some period of time. The geometry of the fastest possible shockwave line dictates that the sections above cannot be impacted by the incident (shown in gray). The sections below this line (shown in blue) could feasibly be impacted by the incident. In actuality, only a subset of the discrete sections (shown in yellow) will be circumscribed by the loading and clearing shockwaves.

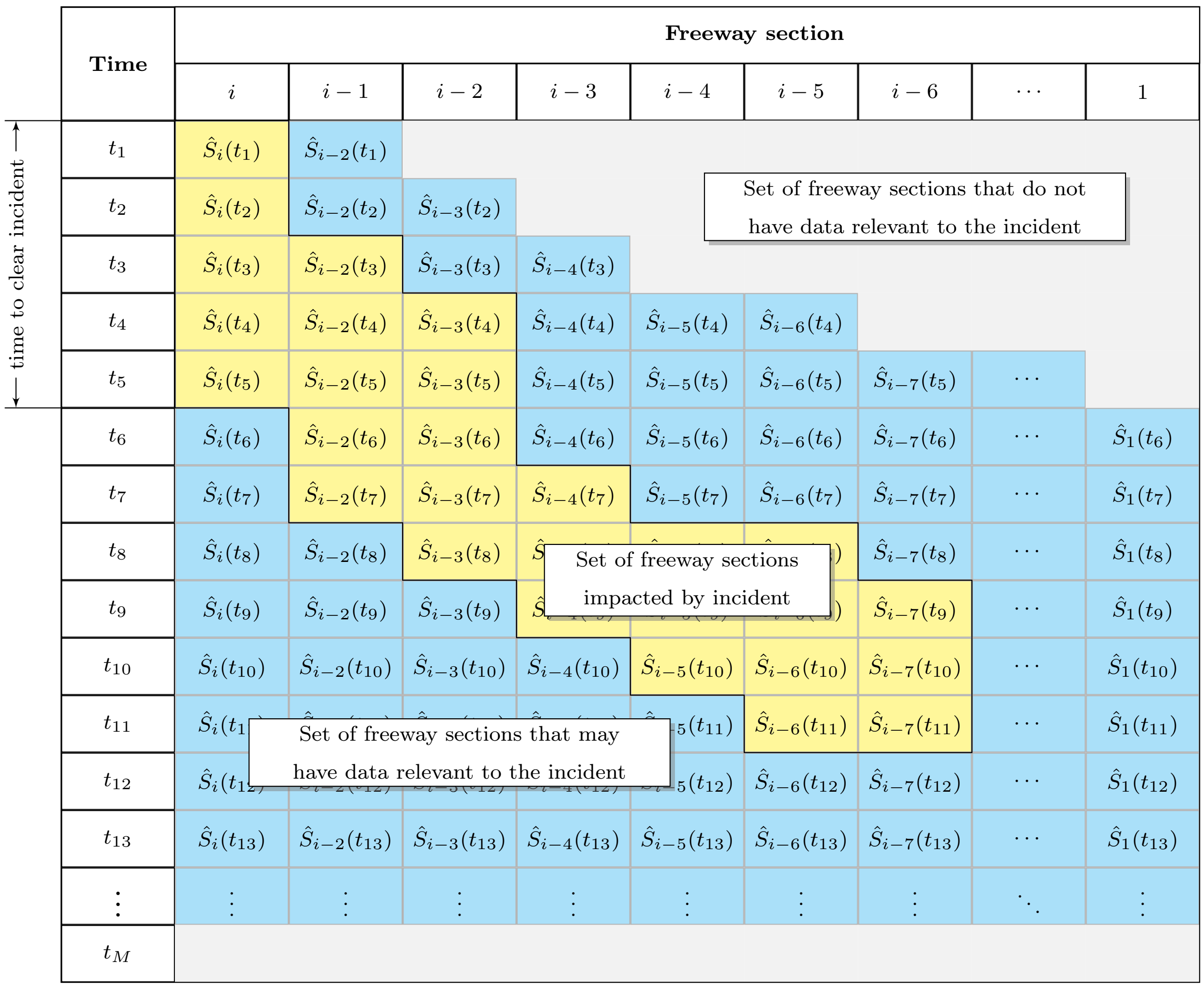

Figure 3.5: Reduced feasible space.

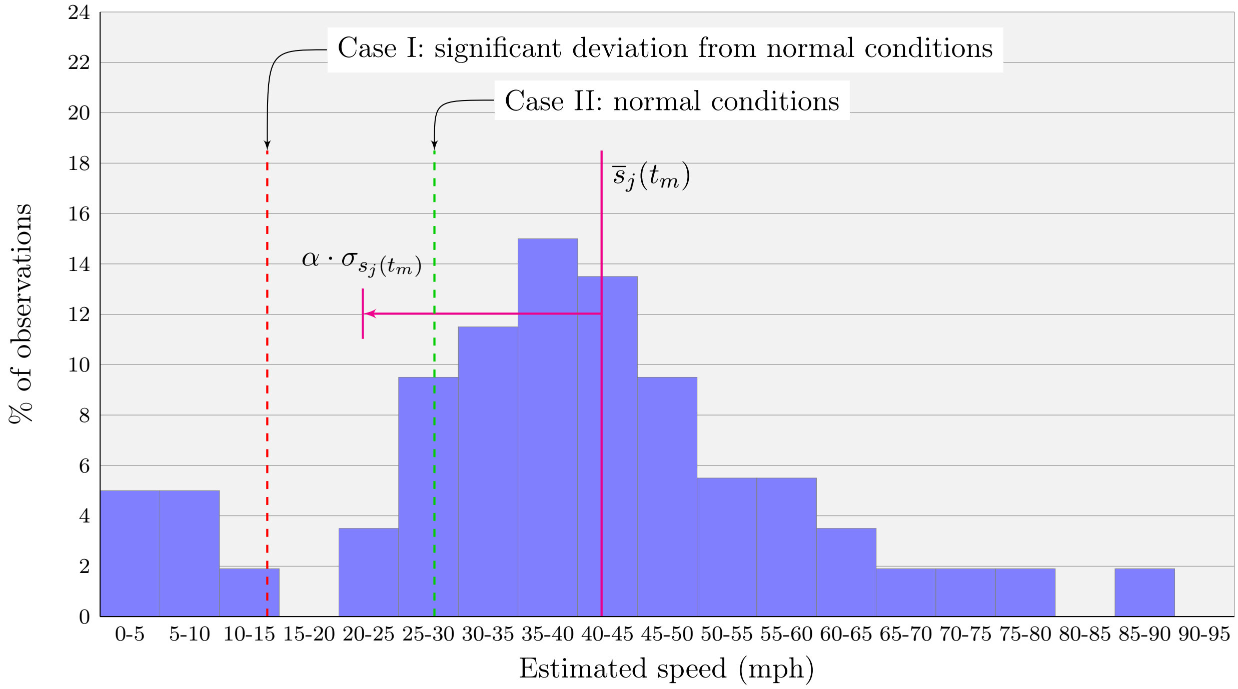

Figure 3.6: The method for determining whether observed speeds are evidence of abnormal conditions. An observed speed distribution for a given freeway section j and time of day m is shown (in blue) having a mean $\overline{s}_j(t_m)$ and standard deviation $\sigma_{s_j(t_m)}$. When the observed speed estimate $\hat{s}_j(m)$ of a section falls below the critical threshold $\hat{s}_j(t_m) \leq \overline{s}_j(t_m) - \alpha \cdot \sigma_{s_j(t_m)}$, the observation is treated as evidence that an incident is occurring. When $\hat{s}_j(t_m) > \overline{s}_j(t_m) - \alpha \cdot \sigma_{s_j(t_m)}$, the observation is treated as evidence of normal conditions.

ble affected sections by an incident of greatest consequence (when all lanes are blocked, as assumed above).

Considering that the speed of traffic in sections adversely affected by the traffic incident will be lowered, the basic idea behind discriminating between these two regions is to compare the incident speed, $\hat{s}_j(t_m)$, to the distribution of the non-incident speeds $s_{jn}(t_m); n = 1, 2, \ldots, n_{obs}; n_{obs} \leq 52$ and assign some level of confidence that any particular $\hat{s}_j(t_m)$ was not drawn from the distribution of $s_{jn}(t_m)$. If we limit our level of confidence that the speed was not drawn from the distribution of $s_{jn}(t_m)$ to those that compute to be at least $\alpha \cdot \sigma_{s_j(t_m)}$, where α is a positive number, below the mean speed of the distribution of $s_{jn}(t_m)$, we can define two cases according the difference between the mean speed and the threshold as shown in figure 3.6. If, in the representation of figure 3.5, we assign to the cells in which speed has been significantly affected by the incident the outcome of the probability that the speed $\hat{s}_j(t_m)$ in such cells is drawn from the distribution of speeds associated with non-incident conditions, we can construct the binary matrix shown in figure 3.7, as the "true"

Time	**Freeway section**								
	i	$i-1$	$i-2$	$i-3$	$i-4$	$i-5$	$i-6$	$\cdots$	1
t_1	0	1							
t_2	0	1	1						
t_3	0	0	1	1					
t_4	0	0	0	1	1	1			
t_5	0	0	0	1	1	1	1	1	
t_6	1	0	0	1	1	1	1	1	1
t_7	1	0	0	0	1	1	1	1	1
t_8	1	1	0	0	0	0	1	1	1
t_9	1	1	1	0	0	0	0	1	1
t_{10}	1	1	1	1	0	0	0	1	1
t_{11}	1	1	1	1	1	0	0	1	1
t_{12}	1	1	1	1	1	1	1	1	1
t_{13}	1	1	1	1	1	1	1	1	1
$\vdots$	1	1	1	1	1	1	1	1	1
t_M									

Figure 3.7: Binary incident speed classification matrix.

representation of the affected region.

However, owing to uncertainties and other externalities, assigning an outcome based on a significance test related to the null hypothesis that the cell speed actually came from the non-incident speed distribution, in general will result in less-than accurate conclusions regarding the affected region, as shown in figure 3.8, where the true boundary of the affected region is shown by the red line.

If we define P_{jm} as follows:

$$P_{jm} = \begin{cases} 0; & \hat{s}_j(t_m) \leq \overline{s}_j(t_m) - \alpha \cdot \sigma_{s_j(t_m)}, \\ 1; & \hat{s}_j(t_m) > \overline{s}_j(t_m) - \alpha \cdot \sigma_{s_j(t_m)}. \end{cases} \tag{3.5}$$

the problem of determining the "best" set of "dot-shaded (or yellow)" cells can be formulated

Time	**Freeway section**								
	i	$i-1$	$i-2$	$i-3$	$i-4$	$i-5$	$i-6$	$\cdots$	1
t_1	0	1							
t_2	0	1	1						
t_3	0	0	1	1					
t_4	0	0	1	1	1	1			
t_5	0	0	0	1	1	1	1	1	
t_6	1	0	0	1	0	0	1	1	1
t_7	1	0	0	0	1	1	0	1	1
t_8	1	1	0	0	0	0	1	1	1
t_9	1	1	0	0	1	0	0	1	1
t_{10}	1	1	0	1	0	0	0	1	1
t_{11}	1	1	1	0	1	0	0	1	1
t_{12}	1	1	1	1	1	1	0	1	1
t_{13}	1	1	1	1	0	1	1	1	0
$\vdots$	1	1	1	1	1	1	0	1	1
t_M									

Figure 3.8: Empirical binary incident speed classification matrix.

as following statement.

$$\sum_{\forall \text{dot shaded cells}} P_{jm} + \sum_{\forall \text{unshaded cells}} (1 - P_{jm}) = \text{Minimum} \tag{3.6}$$

Or, defining

$$\delta_{jm} = \begin{cases} 1, & \text{if cell is affected by the incident (i.e., dot-shaded),} \\ 0, & \text{if cell is not affected by the incident (i.e., plain shaded).} \end{cases}$$

Equation (3.6) can be written as:

$$\sum_{\forall j,m} [P_{jm} \cdot \delta_{jm} + (1 - P_{jm}) \cdot (1 - \delta_{jm})] = \text{Minimum} \tag{3.7}$$

In this minimization problem, care must be taken regarding the number of observations because we use statistics regarding the mean and standard deviation that would be sensitive to the number of observations. It must be noted that a relatively high portion of traffic data cannot be collected due to various reasons, e.g., electrical malfunction of detector and temporary freeway maintenance work. However, when calculating the mean and standard deviation of the distribution of the non-accident speeds $s_{jn}(t_m); n = 1, 2, \ldots, n_{\text{obs}}; n_{\text{obs}} \leq 52$ and assigning some level of confidence that any particular $\hat{s}_j(t_m)$ was not drawn from the distribution of $s_{jn}(t_m)$, we need a "sufficient number" of observations, n_{obs}. We first need to set a threshold regarding the minimum number of observations that we require in order to have some confidence in the statistical calculations for mean and standard deviation, say n_{minobs}. Since 30 is commonly used for the minimum number of observations required for the law of large numbers to apply, we set $n_{\text{minobs}} = 30$. Then, for cells in the matrix in figure 3.7 for which $n_{\text{obs}} < n_{\text{minobs}}$, we set $P_{jm} = 0.5$ and continue the analysis; i.e., we modify the definition of P_{jm} as follows:

$$P_{jm} = \begin{cases} 0; & \hat{s}_j(t_m) \leq \overline{s}_j(t_m) - \alpha \cdot \sigma_{s_j(t_m)}; n_{\text{obs}} \geq n_{\text{minobs}} \\ 1; & \hat{s}_j(t_m) > \overline{s}_j(t_m) - \alpha \cdot \sigma_{s_j(t_m)}; n_{\text{obs}} \geq n_{\text{minobs}} \\ 0.5; & n_{\text{obs}} < n_{\text{minobs}} \end{cases} \tag{3.8}$$

Moreover, the subset of cells for which the accident speeds are significantly different from the non-accident speeds comprise a region that theoretically must obey certain properties. Specifically, there are three impossible local shape configurations for the subset of spatio-temporal cells congested by the accident. The first such case is a region that contains any holes, as shown in figure 3.9. In addition, the vertical position (t) of any dot-shaded ("yellow") section j must be either lower or same (i.e., $\leq$) vertical position of the neighboring shaded ("yellow") section $j - n$, as shown in figure 3.10. Likewise, figure 3.11 represents another impermissible configuration. These conditions can be enforced by the following relationships:

$$\delta_{j+k,m} \leq [1 - (\delta_{j,m} - \delta_{j+1,m})] \cdot R; \forall j, m; \forall k \leq J - j \tag{3.9a}$$

$$\delta_{j,m+r} \leq [1 - (\delta_{j,m} - \delta_{j,m+1})] \cdot R; \forall j, m; \forall r \leq M - m \tag{3.9b}$$

$$\delta_{j,m+k} \leq [1 + (\delta_{j,m} - \delta_{j+1,m})] \cdot R; \forall j, m; \forall k \leq M - m \tag{3.9c}$$

where R is a large number, J is the maximum number of upstream sections, and M is the maximum number of time period, in 5-minute intervals (e.g., if the analysis time period is 4 hours in the study, $M = 48$). Equations (3.7) and (3.9) are in the form of the objective

Time	Freeway section								
	i	$i-1$	$i-2$	$i-3$	$i-4$	$i-5$	$i-6$	$\cdots$	1
t_1	0	1							
t_2	0	1	1						
t_3	0	0	1	1					
t_4	0	0	0	1	1	1			
t_5	0	0	0	1	1	1	1	1	
t_6	1	0	0	1	1	1	1	1	1
t_7	1	0	0	0	1	1	1	1	1
t_8	1	1	0	0	0	0	1	1	1
t_9	1	1	1	0	1	1	0	1	1
t_{10}	1	1	1	1	0	0	0	1	1
t_{11}	1	1	1	1	1	0	0	1	1
t_{12}	1	1	1	1	1	1	1	1	1
t_{13}	1	1	1	1	1	1	1	1	1
$\vdots$	1	1	1	1	1	1	1	1	1
t_M									

Not possible

Figure 3.9: Impossible shape I of the impacted region.

function and constraint, respectively, of Binary Integer Program (BIP) The determination problem described above can be represented in the form of the following BIP problem:

$$
\begin{aligned}
&\min_{\delta_{jm}} Z = \sum_{\forall j,m} [P_{jm} \cdot \delta_{jm} + (1 - P_{jm})(1 - \delta_{jm})] \\
&\text{st} \\
&\delta_{j+k,m} \leq [1 - (\delta_{j,m} - \delta_{j+1,m})] \cdot R; \forall j, m; \forall k \leq J - j \\
&\delta_{j,m+r} \leq [1 - (\delta_{j,m} - \delta_{j,m+1})] \cdot R; \forall j, m; \forall r \leq M - m \\
&\delta_{j,m+k} \leq [1 + (\delta_{j,m} - \delta_{j+1,m})] \cdot R; \forall j, m; \forall k \leq M - m
\end{aligned}
\tag{3.10}
$$

In the definition of P_{jm}, the threshold value α is very critical to separate between speed affected by an incident and the other speeds. In this study, various values were applied to

Time	**Freeway section**								
	i	$i-1$	$i-2$	$i-3$	$i-4$	$i-5$	$i-6$	$\cdots$	1
t_1	0	1							
t_2	0	1	1						
t_3	0	0	1	1					
t_4	0	0	0	1	1	1			
t_5	0	0	1	1	1	1	1	1	
t_6	1	0	0	1	1	1	1	1	1
t_7	1	0	0	0	1	1	1	1	1
t_8	1	1	0	0	0	0	1	1	1
t_9	1	1	1	0	1	0	0	1	1
t_{10}	1	1	1	1	1	0	0	1	1
t_{11}	1	1	1	1	1	0	0	1	1
t_{12}	1	1	1	1	1	1	1	1	1
t_{13}	1	1	1	1	1	1	1	1	1
$\vdots$	1	1	1	1	1	1	1	1	1
t_M									

Not possible

Figure 3.10: Impossible shape II of the impacted region.

find the best threshold. The best value was empirically found from the relation between traffic data and traffic accident as $\alpha = 0.25$.

3.1.5 Improvements to the core delay model

When research is transferred into application, the idealized assumptions underlying the model are often found to oversimplify the real world. In the case of TMCPE, this technology transfer required some fine tuning of the assumptions in a number of areas, which are outlined below.

Time	Freeway section								
	i	$i-1$	$i-2$	$i-3$	$i-4$	$i-5$	$i-6$	$\cdots$	1
t_1	0	1							
t_2	0	1	1						
t_3	0	0	1	1					
t_4	0	0	0	1	1	1			
t_5	0	0	0	1	1	1	1	1	
t_6	1	0	0	1	1	1	1	1	1
t_7	1	0	0	0	1	1	1	1	1
t_8	1	1	0	1	1	1	1	1	1
t_9	1	1	1	1	1	0	0	1	1
t_{10}	1	1	1	1	0	0	0	1	1
t_{11}	1	1	1	1	1	0	0	1	1
t_{12}	1	1	1	1	1	1	1	1	1
t_{13}	1	1	1	1	1	1	1	1	1
$\vdots$	1	1	1	1	1	1	1	1	1
t_M									

Not possible

Figure 3.11: Impossible shape III of the impacted region.

3.1.5.1 Bounding the incident

Computing maximum feasible time-space bounds of an incident impact is critical both to keeping the calculations for each event tractable, as well as to prevent spurious

[**(crindt)** Discuss the use of leading and trailing shockwave constraints]

3.1.5.2 Distance weighting

One common characteristic of real-world data is that incidents do not occur in isolation. Many times these incidents are related, while other times they are independent. Furthermore, a randomly sampled time-space section in the freeway network will often have regions of congestion arising from the random processes underlying traffic flow. As such, it is fairly common for such congestion plumes to be sufficiently close in time and space to a known

Figure 3.12: Random congestion in proximity to a known event

incident locations.

In such cases, a poorly tuned delay algorithm might erroneously attribute the random congestion to the known event. One such case is shown in [**(crindt)** THE FOLLOWING FIGURE]. Here, a known incident causes a congestion plume that is clearly spatially distinct from some random upstream disturbance. Nonetheless the simplified delay calculation model attributes the random congestion to the known event because in the algorithm's objective, it is "cheaper" to incorrectly identify the uncongested are between the plumes as congested rather than miss out on the opportunity to identify upstream congestion as related to the known incident. This bundling of plumes, however, is an artifact of the objective rather than a correct attribution of congestion to the given event.

We considered two approaches to improving the ability of the model to account for such cases. The first approach was to bias the objective function to make false positives more 'costly' to the algorithm than false negatives. More specifically, this altered the objective function in (3.10) to be:

$$\min_{\delta_{jm}} Z = \sum_{\forall j,m} [(1-\beta) \cdot P_{jm} \cdot \delta_{jm} + (1-P_{jm})(1-\delta_{jm})] \tag{3.11}$$

where $\beta \in (0,1)$ is a parameter set to reduce the importance of false negatives. As $\beta \to 1$, the importance of false negatives is reduced. For instance, setting $\beta = 0.1$ effectively makes a false negative 10% less important than a false positive. This method, however tended to be too conservative and would result in the algorithm calculating [very small regions of delay **(crindt)** Give an example of this].

The second method was to alter the objective function to weight the cells based upon their time-space proximity to the estimated location of the incident.

$$\min_{\delta_{jm}} Z = \sum_{\forall j,m} \left(\left|L - \overline{L}\right| \cdot [P_{jm} \cdot \delta_{jm} + (1-P_{jm})(1-\delta_{jm})] \right) \tag{3.12}$$

where $\left|L - \overline{L}\right|$ is the normalized distance in time and space between the cell and the estimated location of the event disruption. By weighting cells that are further from the expected location of the incident higher, this objective emphasizes matching the evidence near the event disruption more than that evidence further away from the incident.

[**(crindt)** Create before and after figure of the importance of this method.]

Figure 3.13: Illustrating the need for a maximum incident speed

We note that the use of this method was made possible by the improvements made to the TMC activity log to add geospatial coordinates from the CHP iCAD system to the incident logs (see section 4.3.1).

3.1.5.3 Maximum incident speed

Empirical analysis of incidents analyzed using the original delay calculation method showed that in some cases the use of the α threshold to identify likely areas of incident-induced congestion would occasionally produce evidence that extended significantly beyond what was apparent to the analyst. An example of this situation is shown in [figure 3.13 **(crindt)** Make this a side by side before/after comparison].

After analyzing a number of incident exhibiting this problem, we noticed that these incidents all tended to have relatively high observed speeds ($\hat{s}_j(t_m) > 65^{\mathrm{mi}}/_{\mathrm{h}}$) and low standard deviations ($\sigma_{s_j(t_m)} < 2^{\mathrm{mi}}/_{\mathrm{h}}$). In these situations, minor fluctuations in observed speeds are tagged as evidence of incident conditions when they likely are not (though the may indicate the imminent onset of delay). Such noise in the evidence results in poor performance for the algorithm.

The obvious solution to this problem was to adjust the α parameter to be greater than 0.25 value determined in earlier research. However, we found that in many cases the problem would only disappear with $\alpha > 3.5$. However, such a high α degraded the method's performance for incidents where $\hat{s}_j(t_m) < 65$.

The best solution we found, however, was to alter the evidence definition to adjust the definition of the evidence P_{jm} to always indicate non-incident conditions if the observed speed $\hat{s}_j(t_m)$ is greater than some *maximum incident speed* threshold $\hat{s}_{max}$. Rewriting equation (3.8), we have:

$$P_{jm} = \begin{cases} 0; & \hat{s}_j(t_m) \leq \overline{s}_j(t_m) - \alpha \cdot \sigma_{s_j(t_m)}; n_{\mathrm{obs}} \geq n_{\mathrm{minobs}}; \hat{s}_j(t_m) < \hat{s}_{max} \\ 1; & \hat{s}_j(t_m) > \overline{s}_j(t_m) - \alpha \cdot \sigma_{s_j(t_m)}; n_{\mathrm{obs}} \geq n_{\mathrm{minobs}} \\ 0.5; & n_{\mathrm{obs}} < n_{\mathrm{minobs}} \end{cases} \tag{3.13}$$

With this new evidence definition, we determined empirically that the combination of $\hat{s}_{max} = 65^{\mathrm{mi}}/_{\mathrm{h}}$ and $\alpha = 1.0$ produced consistent results for a range of incident severities and speeds.

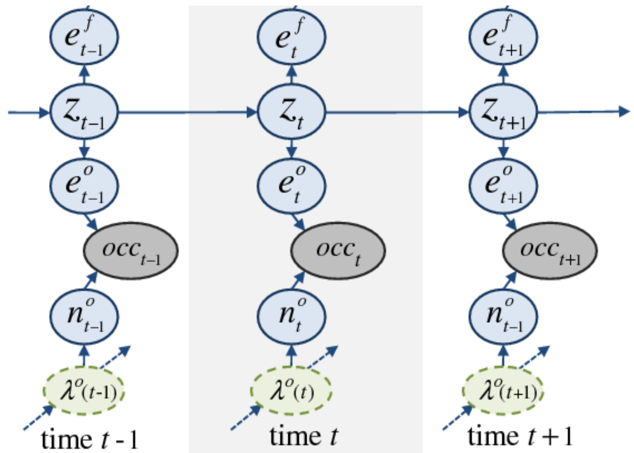

Figure 3.14: Three time segments of the Markov model

3.1.6 Alternative method for computing TMC impacts

The statistical techniques described in section 3.1.4.2 used to determine the evidence that a given time-space freeway segment is experiencing abnormal conditions is admittedly simple. In this section, we present a bottom-up, data-driven approach to modeling traffic behavior from loop-sensor measurements. In section 3.1.7 we describe an unsupervised method that simultaneously filters out sensor failure, learns the underlying normal traffic pattern, and separates and characterizes unusual event activity. Section **??** shows how the output of the model can be used in various applications, including detecting unusual activity, detecting and filtering sensor failure, finding the spatio-temporal signature of accidents, and creating a ranked list of traffic disruptions.

3.1.7 Model Description

Figure 3.14 shows three time-segments of the graphical model for the loop sensor event model. The flow and occupancy processes are joined together via a Markov event process forming two Markov modulated nonhomogeneous Poisson processes[scott00] (MMNHPP). We will discuss each component process in turn.

Figure 3.15: Three time segments of the

Event Process: Modeling Rare, Persistent Events The anomalous measurements that occur during an event caused by a traffic accident or a large concert for example, can be intuitively thought of as being due to a random process changes the observed behavior, increasing or decreasing the measurement for a short period of time. We characterize events as high or low, indicating that the anomalous measurements are larger or smaller than what we would normally expect. In addition, we expect the event type to have continuity; i.e. continuously positive rather than alternating between positive and negative from one time segment to the next. Modeling the event process using a Markov chain, as seen in Figure 3.15, allows us to encode the concept that an event has persistence, but is relatively short in duration and relatively rare and the concept of continuity.

Continuity is achieved in part by defining the event state variable, z_t. In our system we define 8 event states

$$z_t = \begin{cases} 0 & \text{if there is no event at time } t \\ 1 & \text{if there is a high flow, high occupancy event at time } t \\ 2 & \text{if there is a high flow, low occupancy event at time } t \\ 3 & \text{if there is a low flow, low occupancy event at time } t \\ 4 & \text{if there is a low flow, high occupancy event at time } t \\ 5 & \text{if there is a high flow, normal occupancy event at time } t \\ 6 & \text{if there is a normal flow, high occupancy event at time } t \\ 7 & \text{if there is a sensor failure at time } t \end{cases} \tag{3.14}$$

and define the probability distribution over z_t to be Markov in time, with transition probability matrix A

$$A = \begin{pmatrix} a_{00} & a_{01} & \cdots \\ a_{10} & a_{11} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{3.15}$$

where a_{ij} is the probability of transitioning from state i to state j, and where each row of A sums to 1.

The parameters of the transition matrix enable us to build in the concepts of persistence and rarity of events, as well as consistency of events. Persistency is encoded in the event self-transition parameters. For example, the expected value of the length of a high flow, high occupancy event is $1/(1 - a_{11})$. Rarity of events is encoded in the no-event to no-event transition parameter, the expected number of time segments before seeing an event being $1/(1 - a_{00})$. Continuity is managed by making the event self-transition parameters much larger than the event-to-event transition parameters (eg. $a_{11} >> a_{12}, a_{13}$ etc.).

These parameters create "soft" rules that guide the model to recognize events as we understand them. We will show later in Section **??** that these parameters do not

[Failure... event to event transition not unusual, but frequent **(crindt)** wha?]

the number of time-segments between events is geometric with expected value $1/z^{01}$ (in units of time-segments), and the length of each event is geometric with expected value $1/z^{10}$. We give z^{00}, z^{11} priors

$$z^{00} \sim \beta(z^{00}; a_{00}, b_{00}) \qquad\qquad z^{11} \sim \beta(z^{11}; a_{11}, b_{11}) \tag{3.16}$$

where $\beta(\cdot)$ is the Beta distribution, and a_{00}, b_{00}, and a_{11}, b_{11} are parameters for the two Beta priors.

The event process is modeled as an HMM (see Chapter **??**), as illustrated in Figure **??**. Given z_t, we can model the increase in observation counts due to the event, e_t, as Poisson with rate γ

$$e_t \sim \begin{cases} 0 & z_t = 0 \\ \mathrm{P}(e; \gamma) & z_t = 1 \end{cases} \tag{3.17}$$

and γ is independent at each time t

$$\gamma \sim \Gamma(\gamma; a^E, b^E). \tag{3.18}$$

where a and b are parameters of the Gamma distribution and the superscript E indicates the event process.

Just as there are different Poisson rate parameters for different times-of-day in the normal process, one can imagine different Poisson rate parameters for the event process based on time. However, although the observed time and day indicate the appropriate normal process rate parameter to use (e.g. traffic Monday mornings at 11am at a specific sensor location is generally predictable), the same cannot be said for the event process (event traffic on Monday morning at 11am might be due to a small event with relatively little extra traffic or due to a large event with more traffic).

One method of approaching this problem of different sized events would be to define several event types and let the model learn the appropriate rate parameters for each type (perhaps extra traffic due to large sporting events has one distribution and extra traffic due to concerts has a different distribution); however, not only does this increase computational complexity, it could make the model overly flexible and too sensitive to naturally occurring noise in the count measurements (see Section **??** and Chapter **??**). A more robust solution is to marginalize over γ analytically

$$\int \mathrm{P}(e;\gamma)\Gamma(\gamma;a^E,b^E) = \mathrm{NBin}(e;a^E,b^E/(1+b^E)) \tag{3.19}$$

where NBin is the negative binomial distribution with parameters a^E and $b^E/(1+b^E)$.

This negative binomial distribution has a larger variance than the Poisson distribution, and is thus able to model a "catch-all" event type that gives some probability mass to event counts from various-sized event types.

3.2 Modeling TMC impacts

The following details a method for computing transient or oversaturated queues caused by incident conditions subject to critical parameters under the influence of TMC operations—particularly the time to respond to an incident. It represents the basis of the approach we use to model the benefits of the TMC.

We begin by describing the typical capacity-disrupting incident. Let

$A(t)$ = Cumulative quantity (or number of vehicles) to arrive by time t.

$D(t)$ = Cumulative quantity (or number of vehicles) to have been serviced by time t.

$Q(t) = A(t) - D(t)$ = Number of vehicles in the queue (or queue length) at time t.

Suppose that the arrival rate, queue, and time interval are sufficiently large so that random fluctuations in the number of arrivals can be considered small compared to the observed number of arrivals; i.e., the Law of Large Numbers applies:

$$\Pr\left\{\lim_{t\to\infty} \frac{A(t) - E\{A(T)\}}{E\{A(t)\}}\right\} \tag{3.20}$$

Equation (3.20) implies that $A(t), D(t)$ can be approximated by non-random continuous variables.

For example, when $Q(0) = 1$, and we have a steady arrival rate $\lambda \Rightarrow A(t) = \lambda t$, and a

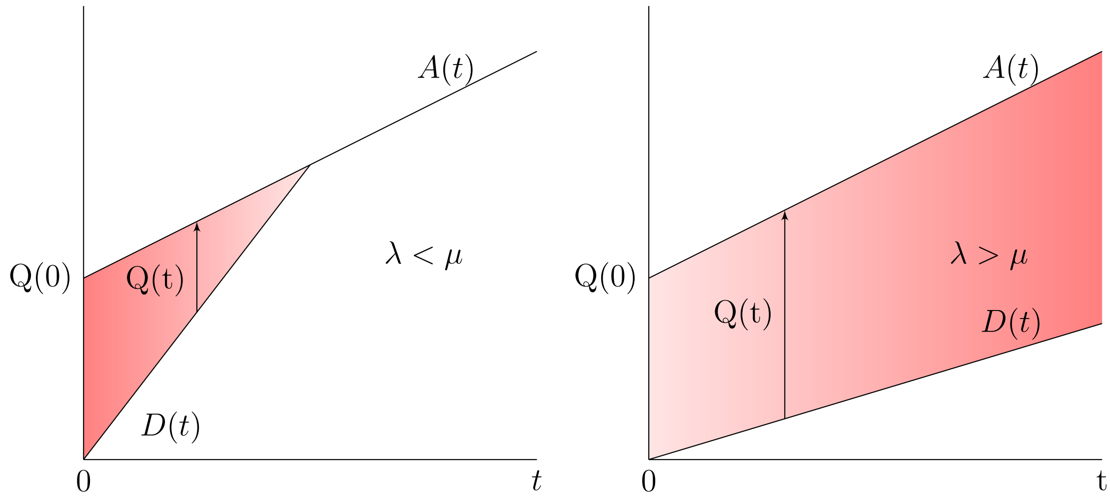

Figure 3.16: Queuing under constant arrival and service rates. In both figures, a queue of length $Q(0)$ exists at time 0 and there is some constant arrival rate $A(t) = \lambda$. The figure on the left shows the dissipation of the queue when the arrival rate is less than the service rate $\lambda < \mu$. The figure on the right shows the increasing queue cased by the oversaturated conditions $\lambda > \mu$.

constant service rate $\mu \Rightarrow D(t) = \mu t, Q(t) > 0$, then,

$$A(t) = Q(0) + \lambda t \Rightarrow \lambda = \frac{dA(t)}{dt}$$
$$Q(t) = Q(0) + (\lambda - \mu)t;\ Q(t) \geq 0$$

These conditions are shown in figure 3.16.

In the general case, the arrival rate λ is not restricted to constant and we have:

$$A(t) = \int \lambda(\tau)d\tau \Rightarrow \lambda(t) = \frac{dA(t)}{dt}$$

Consider flow conditions following an incident that causes the capacity (or service rate $\mu(t)$) to fall below the demand (or arrival rate $\lambda(t)$) for a period during which the incident is identified, verified, and actions are taken to restore capacity to its pre-incident conditions. Assume that the demand remains constant during this period; i.e., $\lambda(t)$ = constant = λ. Prior to the incident, $D(t) \equiv A(t)$. With the onset of the incident, the resulting queue $Q(t)$ can be expected to start off small, rise to a peak (maximum) during the process of clearing (owing to removal activity at the scene), then gradually subside as capacity is restored. Note that, at some time during the clearing process, the rate of growth of the queue will be slowed by actions taken at the scene to begin to restore capacity. Note also that $Q(t)$ will reach its

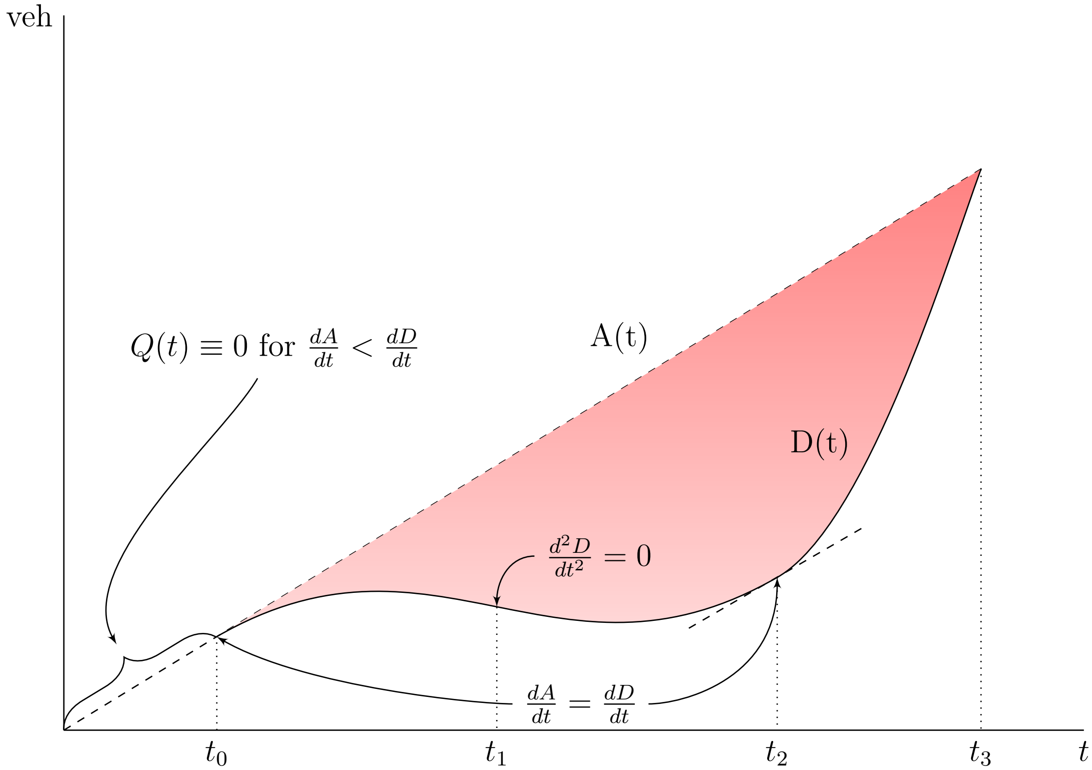

Figure 3.17: The formation of a queuing for a typical incident

maximum precisely when capacity has been restored to a level that is equal to the demand λ; assuming that the full capacity of the section is greater than λ (if, not, the queue will be persistent), clearing of the incident ultimately will permit the queue to be dissipated. This is shown in figure 3.17.

Here, the arrival rate is less than the service rate, i.e., $\lambda < \mu$, until t_0 and the queue is essentially zero, i.e., $Q(t) \equiv 0, 0 \leq t \leq t_0$. Beyond t_0, $\lambda > \mu$ so that the queue grows until t_2—the point at which the rate of service rate $A(t)$ and demand rate $D(t)$ are equal. From t_2 until t_3, the queue diminishes until t_3, when $D(t_3) = A(t_3), t \geq t_3$ as shown in figure 3.18.

From the perspective of TMC operations, the corresponding incident response would look something like figure 3.19.

3.2.1 Computing queues caused by incident congestion

We now turn to the problem of computing delay as a function of the critical time points of incident response that can be influenced by TMC actions.

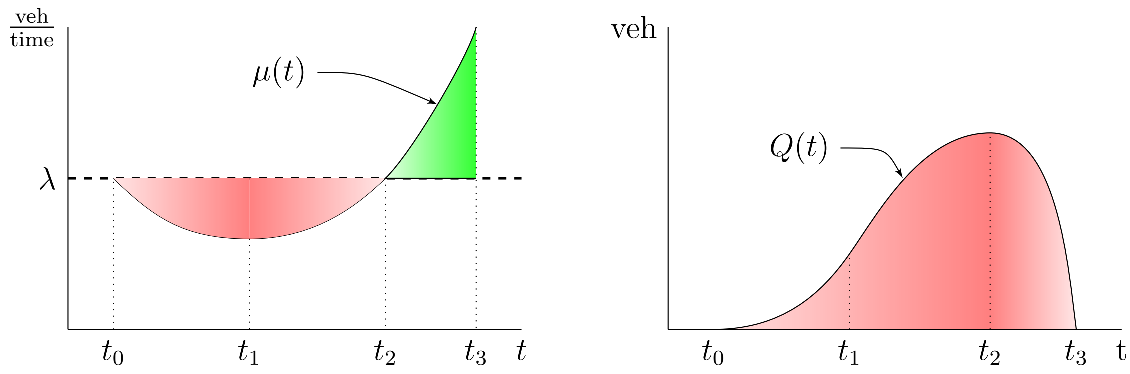

Figure 3.18: The relationship between the arrival rate ($\mu(t)$), the service rate (λ), and queuing ($Q(t)$) during an incident response.

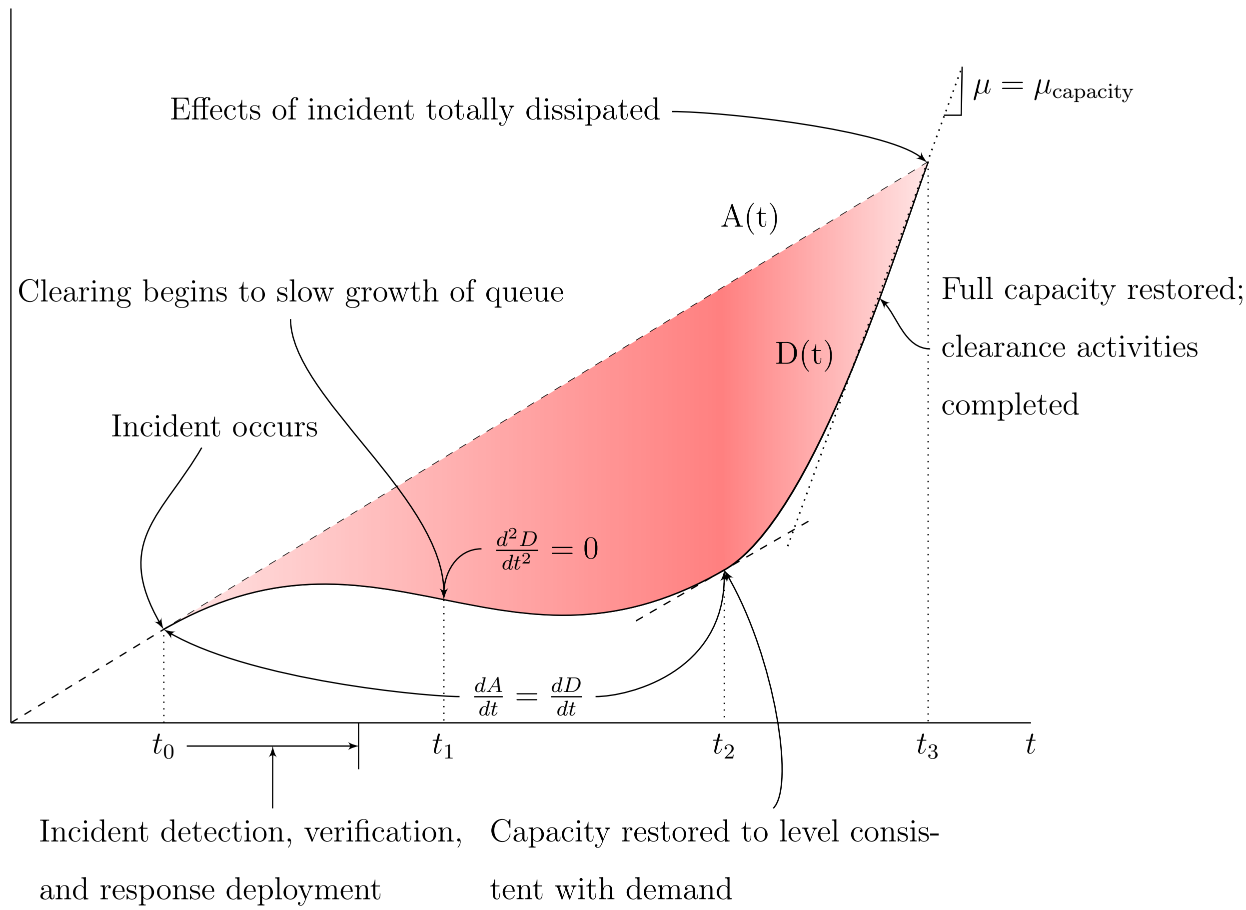

Figure 3.19: TMC incident response and its relationship to queuing dynamics

Case of slight oversaturation

Suppose the time period during which $\lambda > \mu$ is relatively short, i.e., $t_3 - t_0$ is relatively small. Then, expand $\mu(t)$ as a Taylor series about t_1:

$$\mu(t) = \mu(t_1) + \left.\frac{d\mu}{dt}\right|_{t=t_1} \cdot (t - t_1) + \left.\frac{1}{2}\frac{d^2\mu}{dt^2}\right|_{t=t_1} \cdot (t - t_1)^2 + O(t - t_1)^3 \tag{3.21}$$

But, $\frac{d\mu}{dt} = 0 \Rightarrow$

$$\begin{aligned} \mu(t) &= \mu(t_1) + \left.\frac{1}{2}\frac{d^2\mu}{dt^2}\right|_{t=t_1} \cdot (t - t_1)^2 + O(t - t_1)^3 \\ &= \mu(t_1) + \beta(t - t_1)^2 + O(t - t_1)^3, t_0 \le t \le t_3 \end{aligned} \tag{3.22}$$

where

$$\beta = \left.\frac{1}{2}\frac{d^2\mu}{dt^2}\right|_{t=t_1} (\text{Note} : \beta > 0). \tag{3.23}$$

Observe, at $t = t_0$, (3.22) becomes:

$$\mu(t_0) = \mu(t_1) + \beta \cdot (t_0 - t_1)^2 + O(t_0 - t_1)^3 \tag{3.24}$$

Also,

$$\begin{aligned} &\lambda = \lambda(t_0) = \mu(t_0) \Rightarrow \\ &\lambda = \mu(t_0) = \mu(t_1) + \beta \cdot (t_0 - t_1)^2 + O(t_0 - t_1)^3 \end{aligned} \tag{3.25}$$

Similarly

$$\lambda = \lambda(t_2) = \mu(t_2) = \mu(t_1) + \beta \cdot (t_2 - t_1)^2 + O(t_2 - t_1)^3 \tag{3.26}$$

Solving (3.25) for t_0 and (3.26) for t_2, neglecting terms $O(t_2 - t_1)^3$ and $O(t_0 - t_1)^3$ we get

$$\begin{aligned} (t_0 - t_2)^2 = \frac{\lambda - \mu(t_1)}{\beta} &\Rightarrow t_0 - t_1 = \pm \left[\frac{\lambda - \mu(t_1)}{\beta}\right]^{1/2} \\ t_1 > t_0 &\Rightarrow t_0 = t_1 - \left[\frac{\lambda - \mu(t_1)}{\beta}\right]^{1/2} \\ (t_2 - t_1)^2 = \frac{\lambda - \mu(t_1)}{\beta} &\Rightarrow t_2 - t_1 = \pm \left[\frac{\lambda - \mu(t_1)}{\beta}\right]^{1/2} \\ t_2 > t_1 &\Rightarrow t_2 = t_1 + \left[\frac{\lambda - \mu(t_1)}{\beta}\right]^{1/2} \end{aligned} \tag{3.27}$$

Observe from (3.25) and (3.26),

$$(\lambda - \mu(t))|_{t=t_0} = (\lambda - \mu(t))|_{t=t_2} = 0 \tag{3.28}$$

And, from (3.22) we see that $\mu(t)$ is a quadratic in t, so that $\lambda - \mu(t)$ is also a quadratic in t (since λ is a constant). From (3.28), $\lambda - \mu(t)$ has roots $t = t_0, t_2$. So, $\lambda - \mu(t)$ is of the form

$$\lambda - \mu(t) = c(t - t_0)(t - t_2) \tag{3.29}$$

But from (3.23)

$$\beta = \frac{1}{2}\frac{d^2\mu}{dt^2}\bigg|_{t=t_1} \Rightarrow \frac{d^2\mu}{dt^2}\bigg|_{t=t_1} = 2\beta \tag{3.30}$$

Taking d^2/dt^2 of (3.29) and comparing to (3.30), we get

$$\frac{d^2\mu}{dt^2} = 2c \Rightarrow \frac{d^2\mu}{dt^2}\bigg|_{t=t_1} = 2c \Rightarrow c = \beta$$

So

$$\lambda - \mu(t) = \beta(t - t_0)(t - t_2) \tag{3.31}$$

Recall that the queue at any time $t_0 \leq t \leq t_3$ is given by

$$Q(t) = A(t) - D(t) = \int_{t_0}^{t} [\lambda(\tau) - \mu(\tau)]\, d\tau \tag{3.32}$$

For the case of a constant arrival rate, λ, (3.32) becomes

$$Q(t) = A(t) - D(t) = \int_{t_0}^{t} [\lambda - \mu(\tau)]\, d\tau \tag{3.33}$$

Or, from (3.31)

$$Q(t) = \int_{t_0}^{t} [\beta(\tau - t_0)(\tau - t_2)]\, d\tau \tag{3.34}$$

Integrate (3.34) by parts:

$$\begin{aligned} u &= (\tau - t_2) & \qquad dv &= (\tau - t_0)d\tau \\ du &= d\tau & \qquad v &= \frac{1}{2}(\tau - t_0)^2 \end{aligned}$$

$$
\begin{aligned}
Q(t) &= \int_{t_0}^{t} [\beta(\tau - t_0)(\tau - t_2)]\, d\tau = \beta \left\{ \left. \frac{(\tau - t_2)(\tau - t_0)^2}{2} \right|_{t_0}^{t} - \frac{1}{2} \int_{t_0}^{t} (\tau - t_0)^2 d\tau \right\} \\
&= \beta \left\{ \left. \frac{(\tau - t_2)(\tau - t_0)^2}{2} \right|_{t_0}^{t} - \left. \frac{1}{6} (\tau - t_0)^3 \right|_{t_0}^{\tau} \right\} \\
&= \beta \left\{ \frac{(t - t_2)(t - t_0)^2}{2} - \frac{1}{6} (t - t_0)^3 \right\} \\
&= \beta (t - t_0)^2 \left\{ \frac{(t - t_2)}{2} - \frac{(t - t_0)}{6} \right\} \\
&= \beta (t - t_0)^2 \left\{ \frac{(t - t_0)}{3} - \frac{(t_2 - t_0)}{2} \right\}
\end{aligned}
\tag{3.35}
$$

Asymptotic behavior

For $\lim_{t \to t_0+} Q(t)$, we have:

$$
\begin{aligned}
Q(t) = \beta (t - t_0)^2 \left\{ \frac{(t - t_0)}{3} - \frac{(t_2 - t_0)}{2} \right\} \\
\frac{(t_2 - t_0)}{2} \gg \frac{(t - t_0)}{3} \Rightarrow Q(t) \approx \beta (t - t_0)^2 (t_2 - t_0)/2
\end{aligned}
\tag{3.36}
$$

$\lim_{t \to t_0+} Q(t)$*:* At $t = t_2$, $Q(t)$ reaches its maximum:

$$
Q(t_2) = \frac{\beta}{6} (t_2 - t_0)^3 \tag{3.37}
$$

But from (3.27), i.e.,

$$
\begin{aligned}
t_0 = t_1 - \left[\frac{\lambda - \mu(t_1)}{\beta} \right]^{1/2} \\
t_2 = t_1 + \left[\frac{\lambda - \mu(t_1)}{\beta} \right]^{1/2}
\end{aligned}
$$

Then,

$$
(t_2 - t_0) = 2 \left[\frac{\lambda - \mu(t_1)}{\beta} \right]^{1/2} \tag{3.38}
$$

So,

$$
\begin{aligned}
Q(t_2) &= \frac{4}{3}\beta\left[\frac{\lambda-\mu(t_1)}{\beta}\right]^{3/2} \\
&= \frac{4\left[\lambda-\mu(t_1)\right]^{3/2}}{4\beta^{1/2}}
\end{aligned}
\tag{3.39}
$$

Or,

$$
Q(t_2) \propto [\text{Oversaturation}]^{3/2} \tag{3.40}
$$

Note that $Q(t)$ vanishes at $t = t_3$, i.e., $Q(t_3) = 0$. Then, from (3.35)

$$
\begin{aligned}
Q(t_3) = \beta(t_3-t_0)^2\left\{\frac{t_2-t_0}{2}-\frac{t_3-t_0}{3}\right\} &= 0 \Rightarrow \\
\frac{t_2-t_0}{2}-\frac{t_3-t_0}{3} &= 0 \\
t_3 = t_0 + \frac{3}{2}(t_2-t_0)
\end{aligned}
\tag{3.41}
$$

But from (3.27)

$$
t_0 + t_2 = 2t_1 \Rightarrow t_2 = 2t_1 - t_0 \tag{3.42}
$$

Substituting (3.42) into (3.41), we get

$$
t_3 = t_0 + 3(t_1 - t_0) \tag{3.43}
$$

Consider

$$
Q(t) = \beta(t-t_0)^2\left\{\frac{t_2-t_0}{2}-\frac{t-t_0}{3}\right\}
$$

But, from (3.41)

$$
\frac{(t_2-t_0)}{2} = \frac{(t_3-t_0)}{3} \tag{3.44}
$$

Substituting (3.44) into the expression for $Q(t)$,

$$
\begin{aligned}
Q(t) &= \beta(t-t_0)^2\left\{\frac{t_3-t_0}{3}-\frac{t_3-t_0}{3}\right\} \\
&= \beta(t-t_0)^2(t_3-t)/3
\end{aligned}
\tag{3.45}
$$

Relative size of queue

Define

$$
\hat{Q}(t) = \frac{Q(t)}{\max_t Q(t)} = \frac{Q(t)}{Q(t_2)} \tag{3.46}
$$

Let

$$\hat{t} = \frac{(t - t_0)}{t_3 - t_0} \tag{3.47}$$

Then

$$\begin{aligned}\hat{Q} &= \frac{\beta(t - t_0)^2(t_3 - t)/3}{\beta(t_2 - t_0)^3/6} \\ &= 2\frac{\hat{t}^2(t_3 - t_0)^2(t_3 - t)}{(t_2 - t_0)^3}\end{aligned} \tag{3.48}$$

Total delay

Observe that the total delay, D_{total}, is simply

$$D_{\text{total}} = \int_{t_0}^{t_3} Q(\tau)d\tau \tag{3.49}$$

From (3.45)

$$D_{\text{total}} = \frac{\beta}{3}\int_{t_0}^{t_3} (\tau - t_0)(t_3 - \tau)d\tau \tag{3.50}$$

Let $\eta = (t - t_0) \Rightarrow d\eta = d\tau$. Then (3.50) becomes

$$D_{\text{total}} = \frac{\beta}{3}\int_{0}^{t_3 - t_0} \eta^2(t_3 - t_0 - \eta)d\eta \tag{3.51}$$

Let $\phi = \eta/(t_3 - t_0) \Rightarrow d\phi = d\eta/(t_3 - t_0)$. Then (3.51) becomes

$$\begin{aligned}D_{\text{total}} &= \frac{\beta}{3}(t_3 - t_0)^4 \int_0^1 \phi^2(1 - \phi)d\phi \\ &= \frac{\beta}{3}(t_3 - t_0)^4 \left(\frac{\phi^3}{3} - \frac{\phi^4}{4}\right)\Bigg|_0^1 \\ &= \frac{\beta}{3}(t_3 - t_0)^4 \left(\frac{1}{3} - \frac{1}{4}\right) \\ &= \frac{\beta}{36}(t_3 - t_0)^4\end{aligned} \tag{3.52}$$

But from (3.41) and (3.38)

$$t_3 - t_0 = 3(t_2 - t_0)/2 = 3\left[\frac{\lambda - \mu(t_1)}{\beta}\right]^{1/2} \tag{3.53}$$

So, substituting (3.53) into (3.52), we get

$$\begin{aligned} D_{total} &= \frac{\beta}{36}(81)\left[\frac{\lambda - \mu(t_1)}{\beta}\right] \\ &= \frac{9[\lambda - \mu(t_1)]^2}{4\beta} \end{aligned} \tag{3.54}$$

So, total delay is proportional to the square of the amount of oversaturation, $[\lambda - \mu(t1)]$, **or** to the 4$^{\text{th}}$ power of the peak, $(t_3 - t_0)$; i.e.,

$$D_{total} \propto [\lambda - \mu(t_1)]^2 \tag{3.55a}$$

$$D_{total} \propto (t_3 - t_0)^4 \tag{3.55b}$$

3.2.2 Estimating tangible TMC benefits

Equations (3.55) offer a tool for estimating TMC benefits from measurable events in the TMC activity logs. Specifically, if we can determine from TMC logs how much the time to initial response to an incident t_1 is shortened by TMC actions, we can estimate the delay savings attributable to TMC functions. Furthermore, if these actions can be associated with specific processes or technologies, we can make statements regarding the delay savings attributable to specific TMC assets (or bundles of them). The extent to which this is possible depends on the quality of data in the TMC logs, which has been a motivating factor behind the planned improvements to the TMC activity logs discussed in chapter 2.

3.2.3 Ongoing work

[To date, the incident impact model has not been applied to data. The next step in developing this model is to compare baseline delay predictions of the theoretical model (equations 3.55) to the observed delay estimates obtained from the incident impact model (section 3.1). The features of the incident impact model will be determined, in large part, by the improvements to the activity logging system, but we will also make an effort to infer the critical points during past incident management using the legacy logs. **(crindt)** uh]

Chapter 4

System architecture

In this chapter we detail how the TMC incident performance evaluation system has been made available for use in regular TMC operations using a flexible service-oriented architecture that is under active development in the CTMLabs.

4.1 Use Cases

The TMC performance evaluation system could potentially be used in a variety of settings. The primary use-case is to allow TMC personnel to perform post-analysis of incidents that have occurred in order to produce a measure of the value of TMCs to Caltrans and the State of California. In this role, the system must operate as an easy-to-use reporting system with which the analyst can specify a set of analytical parameters defining the incidents to analyze, any necessary modeling assumptions, and TMC components to model. The ideal interface will be accessible from any authenticated computer and will use technology already familiar to operators.

A second potential use-case for the evaluation system is a real-time tmode that attempts to forecast the cost of delaying action on current incidents. While this use is not be part of this initial project, it would be a potentially straightforward extension of the core models to provide real-time decision support to TMC personnel actively engaged in managing multiple simultaneous incidents. In this role, the evaluation system would operate in the background, occasionally broadcasting estimates to other software components for them to integrate into existing TMC processes.

A third use-case is that the performance evaluation system might be integrated into existing or future software systems used by Caltrans and its partners. One obvious example includes the PeMS system, which performs similar functions with less detail. This project is a natural extension to that existing work and may therefore be a candidate for integration.

This software is intended to support a process or set of processes carried out by TMC operators and traffic engineers. Toward that end, the system should integrate as cleanly as possible with existing processes and the software used to support them. We note, however, that existing systems in the Caltrans D12 TMC tend to be monolithic (and generally proprietary) applications—particularly as regards their user interfaces. This makes integrating new software into the TMC very challenging and potentially expensive.

To overcome these barriers, TMCPE system is being deployed using horizontal integration strategies that separate core logic and data from the user interfaces used to interact with them. The initial interface to the system will focus on the post-analysis use-case and will be web-browser based since that is a platform that will be familiar to virtually all users. Integration of the system into any real-time operations will require further study. Nonetheless, the architecture of the system should be flexible enough to allow such deployment in the future.

4.2 Service-Oriented Architecture

The TMC performance evaluation system was being implemented using a service-oriented architecture (SOA) whose logical flows as shown in figure 4.1. A collection of existing databases provide raw data to the Incident Impact Model. An analysis scheduler is configured to regularly execute performance analysis of incidents to generate baseline delays estimates. Analysts can use the TMC impact model via a web interface to analyze the benefits of the TMC for any identified incident.

A possible future extension of the system for real-time delay estimation is not shown in the figure. In this mode, a Real-time Delay Estimator component would actively monitor the system for identified incidents, and would query the incident impact model to estimate the relative costs of delaying action on all active incidents. These results could be broadcast to any TMC management application to assist operators in prioritizing TMC actions. We discuss each of these components in the following sections.

4.3 Databases

4.3.1 TMC activity data

The first interim report for this project (Rindt and Recker, 2008) we noted that

> The D12 TMC activity log records sufficient TMC activity to perform the basic delay calculation and evaluate a limited portion of the benefits provided by the

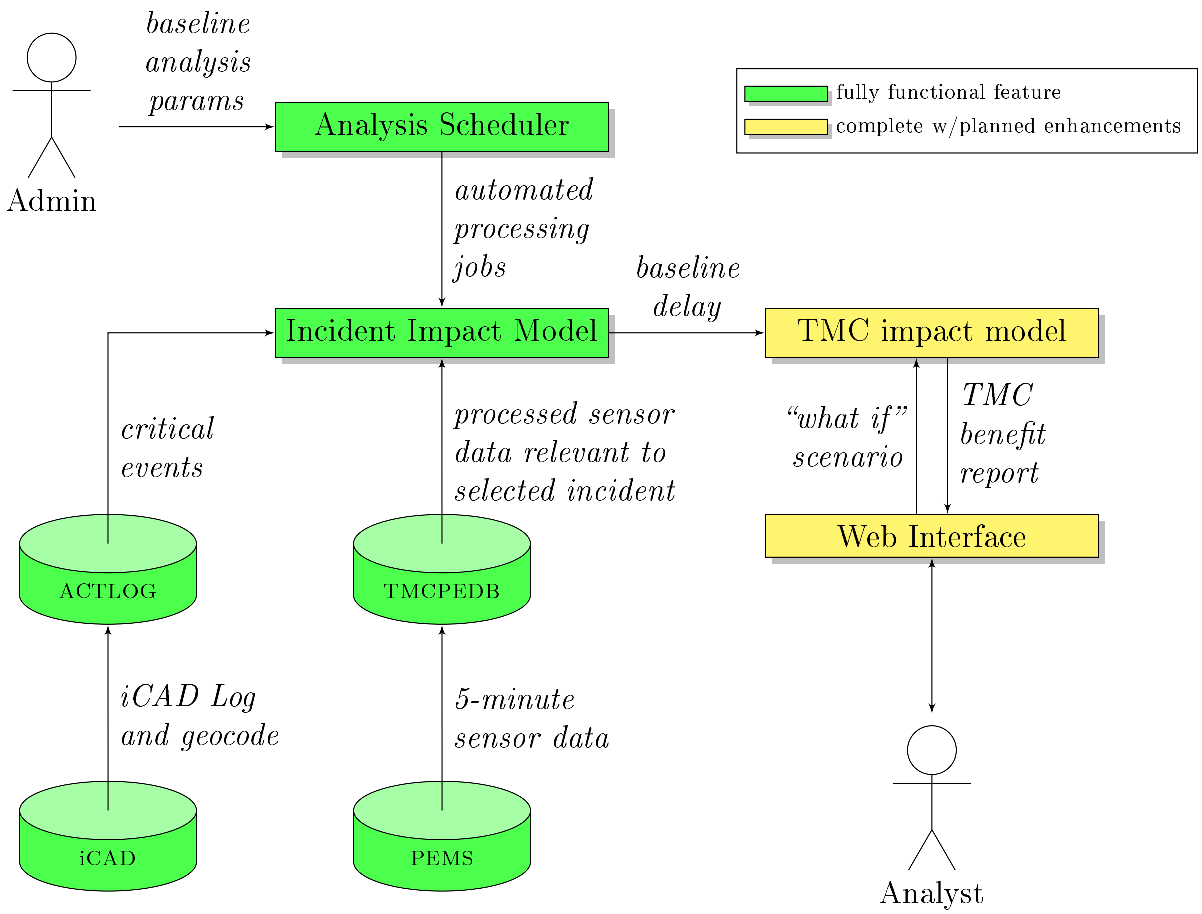

Figure 4.1: The logical system architecture of the TMC performance evaluation system. [**(crindt)** Elaborate]

TMC. For the activity log to be more generally useful to performance evaluation, entries in the log should be tied both to specific technologies that offer notable impacts on TMC processes or to procedures that are designed to have intended effects.

These recommendations were folded into a set of modifications made to the D12 TMC activity log application under UCI subaward #2009-2291 between Caltrans D12 and UCI. The details of this work is included in Appendix A. The most significant changes with respect to the TMCPE analysis is that the activity log was modified to:

- log the specific critical events outlined in chapter 2;
- embed the data from the CHP iCAD log (see section 1.3.2.1) in the activity log database to provide a range of additional event information—most notably a geocoded location

for the incident previously absent from the event dataset; and

- include the data from the TMC Radio Communications logs (RLOG), which offer additional information about resource allocation during incident management.

In order to shield the production database from unnecessary loads from TMCPE users, the development team decided to mirror the TMC activity log database to a secure CTMLabs server. This mirror is further processed during TMCPE analysis to remove any sensitive data such as phone numbers. Consequently, the web application has no access to the raw sensitive data of the source database.

4.3.2 PeMS 5 minute database

The TMCPE service depends on the availability of 5-minute speed and volume data for each section of roadway being analyzed. Though CTMLabs has multiple sources for this data (see section 1.3.1), the team concluded that the PeMS dataset was the most reliable. The CTMLabs deployment therefore mirrors the PeMS Value Added Reseller (VAR) feed, which provides 5-minute section speeds and volumes. These are mirrored onto the CTMLabs dataserver at 5 minute intervals and are loaded into the CTMLabs PeMS database.

In addition to the 5-minute speeds and volumes, the TMCPE delay calculation requires the computation of the expected distribution of flow and speed for affected sections by time-of-day and day-of-week. This is computed using a rolling 52-week horizon as described in equation (3.3). Since this is a relatively expensive computation, it is performed only once on demand the first time an analysis needs it, and then is cached in the database local PeMS database. A saved query, or 'view' is stored in the database that automatically selects the distribution from the cache, if available, or triggers the computation of a new distribution if it is missing. This caching improves the performance of the web application significantly.

4.3.3 TMCPE application database

All domain data generated by the TMCPE application are stored in the application database. This includes all incident analyses and their components (such as individual facility analyses). The details of the domain objects stored in this database are discussed in section 4.4 below.

4.4 Components

The main components of the performance evaluation system fall into two categories: business logic components that process data to produce new metrics for use in the TMC and interface components that expose that data to other software or to users in the TMC.

Currently, the core business logic, including the TMCPE computations is implemented using custom `perl` scripts and the commercial General Algebraic Modeling System (GAMS) optimization software (?). The user interface is implemented using the `Grails` web application framework (?). Both `perl` and `Grails` offer persistence mechanisms for reading and writing data to/from the databases. However, because `Grails` is a convention-oriented framework while `perl` is more of a general purpose toolkit that is easier to adapt, we decided to allow `Grails` to manage the database schemas where they are shared between the business and user interface components. The `Grails` persistence mechanism is built atop the industry standard `Hibernate` persistence (King et al., 2010) engine, and consequently has enterprise stability and scalability.

Thus, the general flow of development is to define the domain classes in the `Grails` framework, allow `Hibernate` to generate (and update) the database schemas as necessary, and then automatically pull the new schemas into the `perl` business logic using the `DBIx::Class` `perl` module. This approach has proven robust and flexible in the development of the system.

4.4.1 Business Logic

4.4.1.1 Incident impact service

This component encapsulates the incident impact model described in section 3.1. It only requires the activity log ID as input and will automatically query the sensor database and CHP iCAD data to obtain further data necessary for the analysis. The service returns a list of impacted sections and associated delay calculations.

The core models for the incident impact service have been developed using a program written in the `perl` scripting language. The high-level logic for this service is shown in figure 4.2. It involves three main steps: obtaining raw event data from all data sources, preprocessing that data to identify the location of the event and the critical incident management steps from the logs, and computing the incident and TMC impacts.

[**(crindt)** Discuss parsing processing, may have to blow up diagram.]

Recommendations for future improvements The current implementation formulates the BIP in equations (3.10) for solution using the ILOG CPLEX ("C"/Simplex) (CPLEX) solver under GAMS. At present, the solution time for determining the impacted region on a single facility for a major incident is on the order of 60 seconds. Since the number of major incidents per day is relatively small, this performance is suitable for batch processes that compute the baseline delays for a given day during the course of the following night. Still, it is likely that the (time) performance of the solver can be improved by using algorithms more tuned for solving BIP problems. Improved performance would give the analyst more flexibility for re-running the solver with different parameters (e.g., α, the maximum duration, etc.) for a given incident. This would, in turn, allow the web-based interface to provide a better end-user experience.

Additionally, the improved statistical techniques discussed in section 3.1.6 have great promise for providing more robust identification of the range of impact for unexpected events in the system. While the current implementation is not sufficiently efficient to be incorporated into the live application as it stands, it could likely be optimized to function with the web application.

4.4.1.2 TMC impact service

The TMC impact service encapsulates the TMC impact model resulting from the work described in section 3.2. This service requires as input the activity log ID of the incident to be analyzed, the output from the incident impact service in order to characterize the critical section and estimate the arrival pattern to the impacted region, and any parameters characterizing changes to the observed incident response As with the incident impact service, this model independently accesses additional databases as necessary to support its analysis. The service updates the TMCPE domain classes for the event to include the estimated delay savings attributable to actions characterized by the input parameters.

[**(crindt)** flesh this out some. Probably just collapse with the above section.]

4.4.2 Interfaces

The core logic generates data objects that can be adapted for the use-cases described in section 4.1. To support these applications, the architecture includes two main interfaces to the core logical services.

4.4.2.1 Web-based incident post-analysis tool

The primary interface to the performance evaluation system is a map-based web interface that allows an analyst to query the model for

- TMC benefits (delay savings) for specific incidents
- TMC benefits for a set of incidents meeting particular criteria (location, time of day, severity, etc.)

The interface will gives analyst control of relevant input parameters to the model, but assumes sensible defaults where possible.

The system uses a web-based interface built atop the grails framework, which is an open-source java-based web application system that is compatible with existing projects in CTMLabs. Within that framework, we are implementing a map-based interface using the OpenLayers `javascript` library (OpenLayers, 2010) and the OpenStreetMap map image tiles (openstreetmap.org, 2010) (Google maps can also be used as a background layer).

Analysis of a particular incident is reflected in the analysis of all potentially impacted facilities, whose estimated time-space impacts are displayed using a custom `javascript` widget developed for this project.

[**(crindt)** Discuss the main features: query screen and its parameters, map view, time-space diagram view, etc.]

4.4.2.2 CTMLabs API

As part of its integration into the CTMLabs architecture, the TMCPE analyses are offered as a using the REpresentational State Transfer (REST) style of exposing services on the Internet (Fielding, 2000) using the CTMLabs API.

The RESTful endpoint is authenticated using the Central Authentication System (CAS) server deployed in the CTMLabs and returns data formatted as Geographic JavaScript Object Notation (GeoJSON) objects (Butler et al., 2008) as shown in figure 4.3[1]. In the TMCPE implementation of this API, the `memo` field is filled with the CAD id of the event in question, the `locString` field with the facility and direction of the affected section, and the `url` field with a link back into the RESTful TMCPE page for the associated analysis.

Using this API, the TMCPE application has been integrated into the CTMLabs project interface, which shows the extent of the TMCPE data available, but does not expose the underlying details to unauthenticated users. [**(crindt)** Project interface picture]

4.4.2.3 Interface security

Access to the data on either of the above interfaces requires that the user logs in to the CTMLabs CAS server to initiate a session. Only those CTMLabs users that have been added to the TMCPE group on the CTMLabs Lightweight Directory Access Protocol (LDAP) server

[1] Technically, the API uses a `javascript` pattern known as JavaScript Object Notation with Padding (JSONP) (Özses and Ergül, 2009) to avoid problems introduced by the *same origin policy* enforced by web browsers. Effectively, however, the API requires the affiliated website—TMCPE in this case—to return a JSON object as described.

are permitted access to the TMCPE data. All interfaces use secure sockets over the HTTP protocol (HTTPS) to encrypt data over the Internet.

4.4.3 Physical deployment of TMCPE Application

The logical system architecture described above (see figure 4.1) is designed to leverage existing software under development for related research being carried out in the CTMLabs. The tasks that were required to deploy the TMC Performance Evaluation platform using this architecture are confined to the specification of the data sources, business logic (software implementing the models described in chapter 3), user interfaces, and information flows between those components.

Figure 4.4 shows how the TMCPE web application was deployed for this project using CTMLabs infrastructure. We discuss this architecture from the perspective of the events that drive the system, namely:

- the availability of new data,
- the processing of new data to produce incident analyses (the business logic), and
- user interaction with the analysis product

4.4.3.1 Incoming Data

As noted earlier, the data for the analysis comes from two main sources: the PeMS 5-minute VAR data and the D12 activity logs (which includes the CHP iCAD data as one component). The TMCPE application is designed to mirror this data.

D12 activity log replication In the case of the D12 activity log, the production `MySQL` database running in the D12 sub network is replicated via a push mechanism to CTMLabs. The mirroring occurs nightly at approximately 2am in order to minimize the impact on the live activity log application in the TMC. The scheduled process uses standard `MySQL` replication methods to push the data to the firewalled CTMLabs data server over the CTMLabs data intertie with D12.

PeMS 5-minute data mirror The structure of the PeMS data mirror differs slightly in that it is a scheduled pull mechanism rather than a push. PeMS offers Value Added Reseller (VAR)accounts to users who can demonstrate a need for the data. To support TMCPE, we established a VAR account for CTMLabs. PeMS publishes new 5-minute summary data for every VDS in the system at roughly 5 minute intervals to an authenticated `ftp` server. The data remains on the system for roughly an hour before it is deleted to make way for new data. To mirror this data, we created a `perl` script to check the PeMS `ftp` site for new data, download it, and push it into a CTMLabs database. This script is scheduled to run every 5 minutes.

4.4.3.2 New incident processing

To provided up-to-date incident analyses and statistics, the D12 activity log data received nightly is processed by `perl` scripts implementing the core business logic. These scripts are currently deployed across two servers. The first machine is used to parse the newly received activity log data to create incident summary objects and scrubbed activity log data in the TMCPE application database.

[**(crindt)** Discuss location parsing. Mentioned in results!!]

Once these objects are created, they are processed through the incident impact and TMCPE impact models, which use the second server to execute analysis jobs—particularly the solving of the mathematical program in equations (3.10) using the GAMS solver. The current implementation uses a synchronous model which pushes the generated mathematical program to the solution server using the `rsync` protocol, executes the GAMS solver on that data via `ssh`, transfers the results back using `rsync`, and processes those results into the TMCPE application domain objects to that they can be served by the TMCPE web application.

4.4.3.3 Web application

The `grails` web application is deployed on a CTMLabs `tomcat` server. The `grails` application is configured to use the TMCPE application database and the PeMS mirror database. We emphasize here that the web application is completely disconnected from the raw activity log database.

Because some of the data provided by the TMCPE application is sensitive, the web application requires that users authenticate using the CTMLabs Central Authentication System (CAS) and its underlying LDAP user database. The system is configured so that only users belonging to the `tmcpe` group on the LDAP server have access to the application and its underlying data. Membership in this group is limited to D12 staff and researchers directly affiliated with the TMCPE project. We will add new users to this group on an as-needed basis based upon guidance given by D12 staff.

4.4.4 Portability of the TMCPE system

4.4.4.1 Licensing restrictions

The TMCPE system was designed from the beginning to be as portable as possible. Where possible, we used open-source off-the-shelf software that runs on multiple architectures and made an effort to select mature products over unproven technology. Table 4.1 summarize the main software components used. Unless noted, all licenses are open source and freely available for Caltrans use (with proper attribution.) Not listed in this table are the CAS and LDAP systems used for user authentication. We assume that any new deployment would need to be integrated with authentications already in place at the site. If not, free and open source CAS and LDAP implementations are readily available for use.

The only obvious major licensing barrier to portability is that the TMCPE system currently uses the commercial GAMS modeling language and CPLEX solver. This is a powerful and expensive modeling system that was used primarily for convenience and speed. As noted in section 4.4.1.1, however, it would be possible to develop a custom (and unencumbered) solver that would perform similarly.

4.4.4.2 Hardware requirements

Beyond the licensing issues, the system requires 1-5 servers for data storage, processing, and for serving the web application. Performance of the system improves when distributed across multiple servers, but nothing precludes deploying everything to a single server.

4.4.4.3 Network requirements

Because the user interface of the TMCPE system is a web application, the server must be on a network accessible to the users. Note that it is easier to secure the raw activity log data if the system is spread across several servers as in the current deployment. If a new deployment wishes to use PeMS data, then a connection to the PeMS `ftp` server is also necessary.

4.4.4.4 Data requirements

As discussed above, the external data required for the TMCPE application include (current sources in parentheses):

- Current and historical 5-minute VDS speed and volume data (PeMS)
- Geocoded TMC activity data with critical events labeled per chapter 2 (D12 activity log, iCAD)

- Map image tiles (for the user interface) (Any tile server compatible with OpenLayers; currently using OpenStreetMap data)

In theory, the TMCPE application is data agnostic as long as the data sources can be mapped onto this input requirements. For the VDS data and the tiled map data, there are likely many available choices that would be suitable. To date, however, we know of no other TMC activity logging application that would meet the activity data requirements. It is probable that Caltrans efforts to develop TMCAL will compatible or could be made compatible with relative ease.

4.4.4.5 Steps to porting the application

The following summarizes the anticipated steps to port the TMCPE application.

1. Confirm availability of VDS, activity log, and map tile data.
2. Obtain deployment server(s).
3. Allocate logical components to physical infrastructure.
4. Install all required software from table 4.1, obtaining the necessary GAMS/CPLEX license for the deployment.
5. Install core TMCPE services and configure for datasources. **NOTE:** ***If the datasources differ from those used in this project, new adaptors would need to be written to bring the data into the analysis framework.***
6. Install TMCPE web application to tomcat and configure for data sources.
7. Run deployment tests.
8. Go live.

1. **Obtain Event Data**: Query activity log database for unanalyzed events since last run

2. **Preprocess Events**: For each unanalyzed event

   (a) Loop over the activity and communications logs and:

   - determine the event type based upon the iCAD ID of the event, which falls into the following broad categories: incidents (accidents, etc.), special events (like sporting events) maintenance/construction activities.
   - import privacy scrubbed data into the application database.
   - identify performance measures information needed to formulate the delay calculation.
   - Use iCAD geocoding (where available) and activity log location strings to identify the approximate time-space location of the event's impact on the system
   - Create a TMCPE event object representing the event and store it in the database

3. **Compute Incident Impact**: For each TMCPE event object

   (a) Using the geocoded location of the approximate event site and the critical incident management events from the activity log, [estimate the maximum time-space bounds **(crindt)** more detail] of the event on primary and secondary facilities

   (b) Query the PeMS mirror database for observed speeds and flows for the maximum time-space impact region.

   (c) Generate the mathematical program that estimates the range of incident impact

   (d) Send the program to the [GAMS service **(crindt)** explain] to be solved.

   (e) Parse the results of the GAMS solution back into the TMCPE domain classes for use by the Application Programming Interfaces (APIs)

Figure 4.2: Incident impact service procedural logic

```
{ "type": "FeatureCollection",
  "features": [
    { "type":           "Feature",
      "id":             <appid>,         // optional
      "geometry":       <geojson geometry object>,
      "properties":
         { "locString": <string describing location>,
           "memo":      <string describing data>,
           "url":       <link back to application url for this object>,
         }
    },
    { "type": "Feature"
      ...
    },
    ...
  ]
}
```

Figure 4.3: CTMLabs API format

[**(crindt)** Insert as-built about here and discuss]

Figure 4.4: Physical architecture of the TMCPE web application

Component	`Function`	`Maturity`	License[a]	`Notes`
`MySQL`	D12 activity log mirror	v5, since 1994	GPL	
`Postgresql`	PeMS mirror and TMCPE app database	v9, since 1986	Postgresql	Chosen for its speed and GIS capabilities
`perl`	Scripting language (and modules) used to process incident data	v5, since 1987	GPL	Perl runs on virtually any platform
`GAMS`	Solving the TMCPE delay computations	v2.22, since 1987	Proprietary	Perpetual license approx $10,000
`Grails`	`java`-based web application framework (and modules)	v1.33, since 2006	Apache	
`Tomcat`	`java` servlet (web application) container	v6, since 1999	Apache	The `Grails` application is deployed to tomcat
`Dojo`	`javascript` user interface toolkit	v1.4, since 2005	Academic Free	
`OpenLayers`	`javascript` map interface toolkit	v2.1, since 2004	Clear BSD	
`OpenStreetMap`	Map tiles in the user interface	since 2005	CC-by-SA	

Table 4.1: Main software components used in the TMCPE application

[a]Except where noted as proprietary, all licenses used in this project are open source.

Chapter 5

Representative results

In this chapter we discuss representative results from the TMCPE system. The system went live as a beta in September of 2010 with incidents processed from January 1^{st}, 2007. The system will continue to process new incident data nightly using the CTMLabs infrastructure. In the following sections, we discuss the characteristics of the data and what they imply for incident performance.

5.1 General statistics

At the time of writing, the TMCPE system had processed a total of $15,135$ events recorded by the D12 TMC between January 1^{st}, 2007 and December 1^{st}, 2010. The general statistics are shown in table 5.1. The events are broken down into various types according to the classification given to the event in the activity log entry.

The most common type are general `incidents`. The common characteristic of these incidents are that their iCAD ID in the system was assigned by the iCAD system rather than in the TMC. This generally means that they are disruptions to the system for which no planning has taken place, which include accidents, medical emergencies, and other capacity reducing events—including most Sigalerts in the database. As such, the severity of this class of event (in terms of average $\mathrm{veh}-\mathrm{hr}$ of delay) is higher than any other.

The remaining analyzed entries include special events at Angel Stadium and the Honda Center—both in Anaheim—which have active pre-planned traffic management. We see lower levels of average delay associated with these events, which are in proportion to their impact on the system (in particular, attendence at Angel Stadium events is typically higher than Honda Center events).

The other major class of events are construction and maintenance events that are similarly planned, but usually occur overnight during periods of low demand and therefore have a

Event type	Analyzed	Total	Analyzed Delay (veh-hr)	Average Delay (veh-hr)
Incident	1,933	5,909	542,936	280.9
Construction	101	6,110	481	4.8
Angel Stadium	30	90	4,682	156.1
Maintenance	24	258	2,146	89.4
Honda Center	16	134	716	44.8
Emergency	4	109	35	8.8
Unknown	29	2,491	3,184	109.8
Total	2,137	15,135	554,179	259.3

Table 5.1: Characteristics of analyzed incidents

relatively small impact on the system in terms of delay.

The `emergency` category comes from a legacy **CAD!** (**CAD!**) ID type that should be classified as a general incident. The final `unknown` category reflects CAD IDs that weren't readily classifiable based upon the information available. Most of these events are likely general incidents as well.

The data shows that roughly 15% of the incidents in the database were analyzed, which is a relatively small number. Breaking it down by type, however, we see that about one third of the general incidents have been processed ($1,933/5,9090 \approx 33\%$) while very few of the construction and maintenance events have been ($\frac{101+24}{6{,}110+258} \approx 2\%$). There are several reasons for these relatively low numbers.

The first is that for the system to perform an analysis, it must be able to identify the incident location. Prior to the middle of 2009, the only source of location information available to the system was to parse the activity log strings to try to identify the route, direction, and location (e.g., nearest ramp) recorded for the incident as discussed in section 4.4.3.2. The general success rate for our parsing algorithm is approximately 50%. Those events whose locations could not be determined were left unprocessed until the algorithm can be fine-tuned or their locations are manually identified.

The second major reason that no processing occurs is if there is no evidence of an impact to the system. This can occur if:

- there is no valid data for the estimated time-space region bounds for the event,
- there is insufficient valid historical data to generate the expected speed distribution for the time-space region bounds for the event, and/or
- no measured speeds in the region indicate a disruption to the system (and therefore there is no impact on the system.)
- range of incident severity

- compare to D35

5.2 Comparison to conventional delay calculation

A major component of the TMCPE application is the

5.3 Computational performance

- Discuss performance/processing, etc

5.4 Sample incident analyses

- Max speed threshold
- Capturing of more delay, not just D35
- Accounting for spectator slowing
- Accounting for network effects (not complete)

Chapter 6

Conclusion

[**(crindt)** Flesh these out; ties to section 1.]

Purpose:

- justify valuable technology, personnel allocations, and maintenance costs,
- identify technologies that aren't meeting their initial promise, and
- identify gaps in current operational strategies that might be filled with new technology deployments.

and, from section 1.2.3:

> The focus in this project has been the development of incident-specific delay-estimation metrics that can be used to evaluate TMC performance.

References

Özses, S. and Ergül, S. (2009). Cross-domain communications with jsonp, part 1: Combine jsonp and jquery to quickly build powerful mashups. Retreived on 12/10/2010.

Al-Deek, H., Garib, A., and Radwan, A. E. (1995). New method for estimating freeway incident congestion. *Transportation Research Record*, 1494:30–39.

Butler, H., Daly, M., Doyle, A., Gillies, S., Schaub, T., and Schmidt, C. (2008). *The GeoJSON Format Specification*.

Choe, T., Skarbodonis, A., and Varaiya, P. (2002). Freeway performance measurement system (PeMS): An operational analysis tool. In *Proceedings of the 81st Annual Meeting*. Transportation Research Board.

Chow, W.-M. (1974). A study of traffic performance models under an incident condition. *Highway Research Record*, 567:31–36.

Erera, A. L., Lawson, T. W., and Daganzo, C. F. (1998). Simple, generalized method for analysis of traffic queue upstream of a bottleneck. *Transportation Research Record*, 1646:132–140.

Fielding, R. T. (2000). *Architectural Styles and the Design of Network-based Software Architectures*. PhD thesis, Department of Information and Computer Science, University of California, Irvine.

Goolsby, M. E. and Smith, W. (1971). Influence of incidents on freeway quality of service. *Highway Research Record*, 349:41–46.

Hall, F. and Persaud, B. (1989). Evaluation of speed estimates made with single-detector data from freeway traffic management systems. *Transportation Research Record*, 1232:9–16.

Heydecker, B. (1994). Incidents and intervention on freeways. Technical report, California PATH. Paper No. UCB-ITS-PRR-94-5.

Jia, Z., Chen, C., Coifman, B., and Varaiya, P. (2001). The PeMS algorithms for accurate, real-time estimates of g-factors and speeds from single-loop detectors. In *2001 IEEE Intelligent Transportation Systems Conference Proceedings*.

King, G., Bauer, C., Andersen, M. R., Bernard, E., Ebersole, S., and Ferentschik, H. (2010). *HIBERNATE - Relational Persistence for Idiomatic Java, version 3.6.0*.

Lawson, T. W., Lovell, D. J., and Daganzo, C. F. (1997). Using input-output diagram to determine spatial and temporal extents of a queue upstream of a bottleneck. *Transportation Research Record*, 1572:140–147.

Lighthill, M. J. and Whitham, G. B. (1955). On kinematic waves. i. flood movement in long rivers. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 229(1178):281–316.

Morales, J. M. (1987). Analytical procedures for estimating freeway traffic congestion. *ITE Journal*, 57(1):45–49.

OpenLayers (2010). *OpenLayers User Guide*. OpenLayers Contributors.

openstreetmap.org (2010). *OpenStreetMap Developer's Guide*.

Recker, W., Chung, Y., and Golob, T. (2005). A tool for the incorporation of non-recurrent congestion costs of freeway accidents in performance management. Technical Report Publication No UCB-ITS-PRR-2005-30, Institute of Transportation Studies, UC Irvine. Accessed: 1 July 2008.

Rindt, C. R. and Recker, W. W. (2008). A measurement-based system for tmc performance evaluation: Interim report and recommendations. Technical report, Institute of Transportation Studies, UC Irvine.

Skabardonis, A. and Geroliminis, N. (2004). Development and application of methodologies to estimate incident impacts. In *Proceedings of the 2nd International Congress on Transportation Research in Greece*, Athens, Greece.

Skabardonis, A., Petty, K., Noeimi, H., Rydzewski, D., and Varaiya, P. (1996). I-880 field experiment: Data-base develepment and incident delay estimation procedures. *Transportation Research Record*, 1554:204–212.

Skabardonis, A., Varaiya, P., and Petty, K. (2003). Measuring recurrent and nonrecurrent traffic congestion. *Transportation Research Record*, 1856:118–124.

TRB (1994). Highway Capacity Manual: Special Report 209.

Wang, Y. and Nihan, N. L. (2003). Can single-loop detectors do the work of dual- loop detectors? *Journal of Transportation Engineering*, 129(2):169–176.

Wirasinghe, S. C. (1978). Determination of traffic delays from shock-wave analysis. *Transportation Research*, 12(5):343–348.

Appendix A

Summary of Work Performed Under UCI Subaward #2009-2291

Summary of Work Performed Under UCI Subaward #2009-2291
TMC Performance Measures Project

Task 1: CMS data Integration Feasibility

The purpose of this task was to determine the feasibility of data exchange between the Activity Log and the ATMS CMS subsystem so that Caltrans can make the decision of whether or not to implement feature.

The purpose of this data exchange is to eliminate duplicate manual entry of the same data into two systems, thus reducing workload, and to reduce data entry errors, thus improving quality.

Two methods could be used to implement a data exchange between the Activity Log and the ATMS CMS subsystem:

1) Manually enter the CMS message and data into the Activity Log then transfer the data to the ATMS CMS subsystem.

2) Manually enter the CMS message into the ATMS then transfer the data to the Activity Log.

Through discussions with the TMC personnel, it was determined that it is standard procedure to first enable CMS messages using the ATMS and then enter the information into the Activity Log later, therefore only the second implementation method was analyzed for feasibility.

CMS integration using the second method consists of implementing the following features in the Activity Log:

1) Provide an indicator on an Activity Log page that indicates that there are CMS messages available to be processed.

2) Clicking the indicator will cause the display of a CMS message selection page. Routine system messages, i.e. Travel Time messages, will not appear.

3) The CMS message selection page will allow any or all CMS messages to be either ignored, marked as rejected, or selected for logging.

4) A CMS message marked as rejected will be removed from the list of CMS messages available to be processed and added to the Activity Log database.

5) A CMS message marked as accepted will cause a page to open to allow the message to be associated with other information in the Activity Log and the appropriate data inserted into the Activity Log database.

The estimated effort required to implement CMS integration between the Activity Log and the ATMS is shown in Table 1.

Table 1

Activity Log pages programming	100 hours
ATMS CMS message interface programming	60 hours
Debugging, training, and support	25 hours
Total	**185 hours**

Task 2: Improve Report Builder

The purpose of this task is to improve the Activity Log’s reporting capability. To that end, additional reports were added to all three of the TMC's activity log applications, i.e. the TMT Responders Log, the Operations Activity Log, and the Communications Radio Activity Log.

The TMT Responders Log had three reports added.

1) A delay calculation report

2) An employee report

3) An EA summary report

All three reports included the ability to run the report for any date range. In addition, the date range subroutine that is used by all reports was improved to be more accurate and error check malformed or invalid dates.

The Operations Activity Log, which is used by the TMC back row, had two reports added and several reports were improved, including the report builder itself.

1) Improvements were made to the report builder to display all activity log columns, including the Performance Measures, Route, Direction, Location, Lanes Blocked and Post Miles columns.

2) Both the Daily Report and the Incident History Report were improved to include data captured from the Communications Radio Activity Log and the CHP iCAD public XML feed.

3) The TMC Operations Activity Report was improved to better capture activity log data and the final report was made so the entries could be edited if needed.

4) A new CHP iCAD Activity Report was added, and access to the new Closure Sheets Report from the Communications Radio Log was added to the Operations Activity Log.

The main focus for this phase of the project was on the Communications Radio Activity Log reports which is used by the maintenance personnel in the TMC front row. As this log was only just completed at the end of last phase, only two reports had been started.

1) All of the reports from the Operations Activity Log were copied to the Communications Radio Activity Log and modified to include the new IMMS field.

2) The report builder was improved to display all activity log columns, including the Performance Measures, Route, Direction, Location, Lanes Blocked and Post Miles columns.

3) The Daily Report and the Incident History Report were improved to include data captured from the Operations Activity Log and the CHP iCAD public XML feed.

4) The Spilled Substance report was improved and the ability to email the report as an attachment was added.

The majority of the work was done on the Closure Sheets Report.

1) The Route and Direction fields were split and both were given drop down lists instead of data entry boxes.

2) Location and Description were split into separate entries and a new Lanes entry box was added.

3) Error checking was added to insure proper formatting of the ID field that corresponds to the CAD field in the Activity Log.

After the work was complete, it was reclassified as an Activity Log Routine and removed from the reports window and a button was added inside the Communications Radio Activity Log to access it.

This was required because the information from the Closure Sheets was now being inserted directly into the Communications Radio Activity Log when the 10-97 field gets entered saving duplication of effort. A new copy of the report was added back to the reports window without the ability to modify the entries.

Task 3: Activity Log Support

From time to time bugs and display errors are discovered and reported by the users of the Activity Log Programs. Once they are reported, they are researched and programming fixes are implemented.

The following items were the major fixes or changes in this phase of the contract:

1) Fixed a bug that would cause the JavaScript to break when illegal characters were used in CAD numbers.

2) Fixed the bug that was preventing the transactions with no CAD number from showing up in the Find Record Search.

3) Set the default display when you pick a date to show records with no CAD number.

4) Fixed a bug that caused the menus to display poorly when the screen resolution was set below 1024x768.

5) Fixed a display problem that caused older CAD numbers to show before newer ones.

6) Resolved an issue causing the time to not input properly under some circumstances.

7) Fixed the Communications Radio Log Layout in Internet Expolorer.

8) Fixed a display problem with low-resolution settings on Communications Radio Log.

9) Fixed a problem with duplicate menu items in Communications Radio Log.

10) Diagnosed and fixed CHP iCAD XML data that stopped working on 3/9/2010.

11) Worked on numerous small bug fixes, display fixes, maintenance issues and program changes.

Task 4: Event Management Data Integration Feasibility

The purpose of this task was to determine the feasibility of event/closure/incident data exchange between the Activity Log and the ATMS Event Management subsystem so that Caltrans can make the decision of whether or not to implement feature.

The purpose of this data exchange is to eliminate duplicate manual entry of the same data into two systems, reduce workload, eliminate data entry errors, and thus improve quality.

Through discussions with the TMC personnel, it was determined that the Activity Log should be the source of event data and the method which most conforms to existing procedures would be to transfer data, e.g. type, location, duration, etc., from the Activity Log to the ATMS Event Management subsystem. Once the data has been transferred to the Event Management subsystem, all of it capabilities, such as response plan generation, may optionally be used to further manage the event.

Event Management integration consists of implementing the following features in the Activity Log:

1) A button will be provided on an Activity Log page to initiate the transfer of information from the Activity Log to the ATMS Event Management subsystem.

2) Clicking the initiate button will cause the display of a page to appear which will allow the information to be reviewed by an operator before transfer and to allow for entry of items that are required by Event Management, but are not available in the Activity Log, by the use of pull-downs or clickable lists. The operator may either confirm the data as a valid incident/event/closure and send it to the Event Management subsystem or cancel the entire operation.

3) The transfer of information to the Event Management subsystem will cause the creation of a termination button to appear on the Activity Log.

4) Clicking the termination button will cause the display of a page to appear which will allow the event to be terminated within the Event Management subsystem, thus removing the incident as if it were terminated using the ATMS GUI.

The estimated effort required to implement event management integration between the Activity Log and the ATMS is shown in Table 2.

Table 2

Activity Log pages programming	80 hours
Event Management Oracle database analysis	20 hours
Oracle interface programming	20 hours
Debugging, training, and support	25 hours
Total	**145 hours**

Task 5: Activity Log Enhancements

The purpose of this task was to add improvements to the existing Activity Log Program. As Operators use the program, they discover items that can be improved or items that might have been missed in initial development.

One of the first items was to change the layout of the Communications Radio Activity Log. The columns were moved, the buttons were moved to a more central location, and new buttons were added for closing CAD numbers and IMMS numbers. More room was made available for the memo field, and the bottom selection box was widened so more information can be displayed.

Other changes include:

1) Mouse over menu items were added.
2) Operators were provided with the ability to edit the mouse over menu items.
3) The contact information program was updated.
4) A button was added to allow users to change their password.
5) The (Contacted) check box column for the log was removed.
6) The edit record function was updated to open an incident if a CAD was added while editing a record.
7) Error checking was added to manually entered date stamps on CAD numbers.
8) New search for IMMS number capability was added to Incident History Search.
9) Users can now select the number of records that appear in the lower search box.
10) Several other small miscellaneous changes and updates were made.
11) The Bugzilla bug reporting system was implemented on an available District 12 server and a bug report button linking to that system was added for all logs.

Project cost summary

Table 3 shows the original estimate for the hours required for each task, the actual hours used, and any remaining hours that are unused.

Table 3

Task	Estimated hours	Used hours	Remaining hours
CMS data Integration Feasibility	50	50	0
Improve Report Builder	200	200	0
Activity Log Support	100	100	0
Event Management Data Integration Feasibility	50	50	0
Activity Log Enhancements	100	100	0
Total	500	500	0

Appendix B

Acronyms

ATMS	Advanced Transportation Management System
TMC	Transportation Management Center
PeMS	Performance Measurement System
TMCPE	TMC Performance Evaluation
CAS	Central Authentication System
CTMLabs	California Traffic Management Laboratories
LDAP	Lightweight Directory Access Protocol
API	Application Programming Interface
CHP	California Highway Patrol
iCAD	Intelligent Computer Aided Dispatch
CMS	Changeable Message Sign
RMS	Ramp Metering Station
HAR	Highway Advisory Radio
UCI	University of California, Irvine
D12	Caltrans, District 12
TMCAL	Traffic Management Center Activity Logging

TMT	Traffic Management Team
HAZMAT	HAZardoes MATerial
REST	REpresentational State Transfer
GeoJSON	Geographic JavaScript Object Notation
JSONP	JavaScript Object Notation with Padding
BIP	Binary Integer Program
GAMS	General Algebraic Modeling System
CPLEX	ILOG CPLEX ("C"/Simplex)
GIS	Graphical Information System
VDS	Vehicle Detector Stations
VAR	Value Added Reseller