TOWARDS AN AUTONOMIC ARCHITECTURE FOR REAL-TIME TRAFFIC NETWORK MANAGEMENT

Hamideh Etemadnia, M.Sc.
Department of Civil and Environmental Engineering
Bobby B. Lyle School of Engineering
Southern Methodist University
PO Box 750340
Dallas, TX, 75275-0340
Tel: (214) 768 3024
Fax: (214) 768 2164
hetemadnia@mail.smu.edu

Khaled Abdelghany, Ph.D.
(Corresponding Author)
Department of Civil and Environmental Engineering
Bobby B. Lyle School of Engineering
Southern Methodist University
PO Box 750340
Dallas, TX, 75275-0340
Tel: (214) 768 4309
Fax: (214) 768 2164
khaled@lyle.smu.edu

Salim Hariri, Ph.D.
Department of Electrical and Computer Engineering
University of Arizona
PO Box 210104
Tucson, Arizona, 85721-0104
Tel: (520) 621-4378
Fax: (520) 621-8076
hariri@ece.arizona.edu

March, 2011
ABSTRACT
This paper presents an autonomic-based architecture for real-time traffic management in congested urban transportation networks. The architecture assumes the availability of a spatially distributed set of local controllers in the network. Each controller is capable of monitoring the traffic within a predefined subnetwork and provides efficient control strategies for its traffic. Following a predefined communication protocol, controllers are assumed to be able to share information on the observed traffic pattern and their control actions. In addition, controllers could be dynamically configured to operate in teams to develop integrated traffic management schemes that best cope with the observed traffic pattern in the network. This paper presents the results of an off-line testing of the proposed architecture. It investigates the most efficient team-formation strategies among controllers to mitigate a non-recurrent traffic congestion situation in a typical highway network. The results show that more efficient traffic management strategies could be obtained through collaboration among individual controllers, which result in considerable travel time savings.

KEYWORDS:
Traffic Network Management, Autonomic, Dynamic Traffic Assignment, Distributed Systems

BACKGROUND
Metropolitan areas in the U.S. and abroad are suffering severe traffic congestion problems that are depicted by considerable travel delays, poor environmental quality, high accident rates, and increasing traveler frustration (Schrank & Lomax 2007). With the limited ability to expand the physical capacity of the existing transportation system to meet the growing demand, there are increasing calls for the development of innovative technologies to improve the system’s throughput, efficiency and safety. Intelligent Transportation Systems (ITS) have emerged over the last few decades as one plausible approach. They integrate advances in communications, surveillance, and computational technologies to provide real-time traffic management capabilities for congested urban networks. Thus, considerable investment has been allocated to furnish the transportation infrastructure with hardware and software technologies that support various ITS services. In addition, Traffic Management Centers (TMCs) have been established in many cities to oversee the use of these technologies to alleviate traffic congestion. Nonetheless, traffic network management efforts at most TMCs have been limited to using traffic surveillance technologies to monitor network conditions and to warn of any detected irregularities (e.g., roadway incidents). In most cases, network managers rely on their experience to evaluate the impact on network performance and to decide if a traffic management scheme can be deployed. A scheme is typically restricted to warning travelers of delays associated with an incident, and providing them with primitive route diversion recommendations that most likely do not guarantee the optimal use of the available network capacity.

Realizing the limitations of current practice, considerable research effort has been devoted over the last two decades to studying the problem of real-time traffic network management with the goal of providing real-time decision support capabilities by integrating available ITS services including Advanced Traveler Information Systems (ATIS), Advanced Traffic Management Systems (ATMS), and Advanced Public Transportation Systems (ATPS) (Ben-Akiva et al. 1997a, 1997b; Cascetta et al. 1993; Cuena et al. 1995; Hawas 2004; Head &

As illustrated in Figure 1, three main architectures have been envisioned to develop a real-time traffic management system: centralized, decentralized, and hybrid. The centralized architecture assumes the existence of one single controller that maintains perfect information on the current state of the entire network. This controller is capable of predicting future network conditions, and developing an optimal traffic management scheme including normative route guidance instructions and optimal signal timings (Peeta & Mahmassani 1995a, 1995b; Abdelghany et al. 1999). The decentralized architecture assumes that the network is controlled by multiple controllers (Bazan 1995; Hawas & Mahmassani 1996; Hernandez et al. 2002; Logi & Ritchie 2002; Pavlis & Papageorgiou 1999). Each controller manages traffic within a pre-defined subarea. The route guidance instructions and other control strategies are generated by each controller using real-time traffic surveillance data for the portion of the network within the controller’s locality. The controller guesses the network conditions outside its boundaries. The hybrid architecture assumes that the management scheme is generated using one centralized controller and multiple decentralized controllers with different levels of coordination, if any (Chiu & Mahmassani 2002; Choy et al. 2003; Cuena et al. 1995).

Several drawbacks are reported for these architectures (Peeta & Ziliaskopoulos 2001). For example, the centralized architecture requires predicting state conditions and solving normative route guidance optimization problem for the entire network. These two tasks require intensive computational resources that would likely preclude the centralized architecture from real-time applications, especially for large-scale networks. Also, as information on the current state of the network becomes unavailable to the centralized controller, the robustness of the developed traffic management scheme weakens significantly. Although the decentralized architecture is more likely to meet the real-time execution requirements, it is suboptimal by definition as each controller optimizes the traffic operation within its boundaries with limited/no knowledge of traffic conditions in other parts of the network. Finally, the hybrid architecture requires activating the centralized and decentralized controllers simultaneously. In addition to the difficulty associated with running the centralized controller in real-time, current research work has not developed the ability for elaborate collaboration between the controllers, which limits the potential benefits that could be achieved from their integration.

This paper presents an autonomic-based architecture for real-time traffic management for congested urban transportation networks. The autonomic paradigm is inspired by the human autonomic nervous system that handles complexity and uncertainties, and aims at realizing
dynamic systems and applications capable of managing themselves with minimum human intervention. The architecture assumes the availability of a spatially distributed set of local controllers (autonomic managers) in the network. Each controller is capable of monitoring the traffic within a predefined subnetwork and provides efficient control strategies for its traffic. Following a predefined communication protocol, controllers are assumed to be able to share information on the observed traffic pattern and their control actions. In addition, controllers could be dynamically configured to operate in teams to develop integrated traffic management schemes that best cope with the observed traffic pattern in the network. This paper presents the results of an off-line testing of the proposed architecture. In particular, it investigates the most efficient demand data sharing and team-formation strategies among controllers to mitigate a non-recurrent traffic congestion situation in a transportation network that represents a part of the Dallas Metroplex. Using a dynamic traffic assignment simulation platform, the network performance under rerouting strategies that are produced by the different controller teams is compared. The results illustrate that more efficient routing strategies could be obtained through collaboration among individual controllers, which result in considerable travel time savings during the incident. This paper is organized as follows. The next section describes the autonomic control paradigm and its application to traffic network management. A formal definition of the problem is then given. Next, the role of the local controllers and their demand data sharing mechanism are described. The methodology for coalition structuring among local controllers is then described. The results of a set of off-line experiments that illustrate the network performance under different team-formation strategies among the local controllers are then provided. Finally, conclusions and extensions of this research work are given.

THE AUTONOMIC ARCHITECTURE

To the authors’ best knowledge, this paper is among the first attempts that propose adopting the autonomic paradigm for real-time traffic network management. The concept has been recently introduced to manage large-scale distributed computing infrastructure (Ganek & Corbi 2003; Kephart & Chess 2003; Sterritt et al. 2005; Tesauro et al. 2004). The autonomic architecture models the computing infrastructure system in the form of multiple interacting hardware and software elements, termed autonomic elements (e.g., application server, database, web server, data adaptor, mobile device, printer, etc.). Each element is equipped with an autonomic manager that continuously monitors the element’s current state as well as its external environment, analyzes its status, plans for any needed actions, and executes these actions. In addition, the architecture facilitates collaboration among the different autonomic managers in order to maximize the system’s overall performance.

As illustrated in Figure 2, the proposed autonomic architecture for traffic network management systems adopts a dynamic hierarchal mechanism. The architecture is defined in terms of a set of local controllers (lower-level autonomic managers) that maintains real-time knowledge of their subnetworks. An autonomic manager includes monitoring, decision, and effect modules. Thus, these local controllers perform all sensing activities for the system and are also capable of individually reacting, through their decision module, to the local traffic conditions. In addition, controllers could be dynamically configured to coordinate their actions (i.e., integrate their decision modules) forming a team (upper-level autonomic manager). The figure illustrates an instance of a configuration to manage an observed traffic pattern in a hypothetical network. In this configuration, eight local controllers are activated. The configuration includes two teams as illustrated by their members’ joined subnetworks. One team includes four controllers, and the
other team includes two controllers. As shown in the figure, the remaining two controllers are assumed to operate independently.

The proposed traffic network management system emulates the human’s autonomic behavior in several aspects. First, the system closely represents the structured multilevel behavior that is observed for humans including reflection, routine, and reaction (Norman et al. 2003). Local controllers are the engine for the system’s reaction and routine behavior. The upper-level controllers could involve a greater depth in terms of information processing and development of management schemes. Second, the system replicates the human body in selectively activating subsystems/organs that are most suited for performing a given task. If part of the network is subject to any irregularity, the system is designed to activate local controllers that are capable of collaborating to recover from this detected irregularity. In addition, as the human body is designed to coordinate the actions of its organs, the proposed system ensures the synchronization of the actions of the different controllers to ensure the generation of an efficient traffic management scheme. Finally, as the human body continuously changes its response to cope with the dynamic nature of the external environment, the system is set to modify its configuration to maintain the highest possible performance. The activated set of local and upper-level controllers dynamically changes based on the encountered traffic conditions.

**PROBLEM DEFINITION**

Given is a roadway network \( G(N,A) \), where \( N \) is the set of nodes and \( A \) is the set of links. The horizon of interest is divided into \( T \) departure intervals and \( T' \) observation intervals. Traffic originates from origins \( i \in N \) to destinations \( j \in N \) during the different departure time intervals. We use \( r_{ij}^T \) to represent the number of vehicle trips that originate from origin \( i \in I \) to destination
During departure interval $\tau \in T$. These trips use a set of superior routes $K$ such that $r_{ijk}$ is the number of trips between origin-destination pair $i - j$ in departure interval $\tau$ that use route $k$. A set of non-overlapping distributed controllers $C$ are assumed to cover the entire network such that each controller covers a subnetwork $G^c(N^e, A^e)$, where $N^e$ and $A^e$ are the set of nodes and links in the vicinity of controller $c \in C$, respectively.

Through adequate traffic surveillance, each controller receives real-time updates on the main traffic state variables within its subnetworks at some predefined observation interval $t \in T'$. In addition, historical information on the traffic distribution pattern in the subnetwork of each controller is assumed and is given in the form of time-dependent link-flow proportions (also known as the assignment matrix). These link proportion matrices provide information on how the demand from the different origin-destination pairs is contributing to the observed flow on a link at any observation interval. Each controller is assumed to be capable of estimating the time-dependent origin-destination trip demand as function of the observed traffic and available historical information. The demand within the subarea of controller $c \in C$ is represented by $r_{ij}^{tc}$, where $\{i,j\} \in N^c$, $\tau \in T$, and $N^c \in N$, respectively. This demand represents the local demand that originates and ends within the subarea as well as the demand crossing any of the subarea’s boundaries. For example, if a trip passes through a subnetwork, this trip is counted as part of the demand of this subnetwork. Considering the sub-route of a trip within the subnetwork, the trip is considered to originate from the first node in this sub-route, and ends at the last node in the sub-route. Define $y_{at}$ as the observed number of vehicles on link $a \in A$ during interval $t \in T'$. Also, define $p_{at}^{c-ij}$ as the link proportion for all linksobservation intervals in the subnetwork of controller $c$. The superscript $c$ is used to indicate that these link proportions are computed with respect to the demand of subnetwork $c$. Assuming the random error term is $\epsilon_{at}$, the measurement equation that relates the demand $r_{ij}^{tc}$ to the traffic count observation and the link flow historical information is given in (1) (Cascetta et al. 1993).

$$y_{at} = \sum_{\tau} \sum_{i} \sum_{j} p_{at}^{c-ij} r_{ij}^{x} + \epsilon_{at} \quad \forall a, t$$ (1)

As part of the different control strategies that can be implemented by a controller, each controller can guide the traffic within its boundaries following a predefined routing strategy (e.g., descriptive or normative). For instance, if the initial (historical) route of a vehicle is represented by the node string $\{i, j, k, l, m, n, o, p, q\}$, and this vehicle’s sub-route $\{l, m, n\}$ is part of the subnetwork of controller $c \in C$ (i.e., $\{l, m, n\} \in N^c$), this vehicle could be rerouted, if needed, to sub-route $\{l', m', n'\}$, where $\{l', m', n'\} \in N^c$. Thus, the new route can be represented by the node string $\{i, j, k, l', m', n', o, p, q\}$. In addition, given the current network conditions, a subset of these controllers could be dynamically configured to form one or more teams. Controllers in one team are assumed to share information on their local demand patterns. The team is assumed to estimate the demand for its joined subnetwork. For example, if subarea $m$ is resulting from joining the subareas of controllers $\{c, c', c''\} \in C$, we use $d_{ij}^{mt}$ to define the (merged) demand between the different origin-destination pairs in subarea $m$, where $\{i, j\} \in N^m$.

In this context, two interdependent questions naturally arise: a) what are the best teams of controllers that should be formed at any given time, if any? and b) what are the optimal control/rerouting strategies that individual controllers and any formed teams of controllers should adopt? Providing answers to these two questions is the ultimate goal of the autonomic traffic management architecture described above. Nonetheless, one can recognize the difficulty
of this combinatorial problem. The large problem size, in terms of team possibilities and the controllers’ adopted strategies, requires equipping these local controllers with an array of self-management capabilities. Such capabilities allow the controllers to anticipate the traffic conditions and determine the best control strategy to be adopted either individually or as part of a team. This paper represents a step towards achieving these self-management capabilities. It partially answers the above two questions by studying efficient demand data sharing and team configuration (coalition structuring) among the local controllers. More details on these two problems are given in the next two sections.

**DTA Capabilities and Demand Estimation**

In this research work, controllers are assumed to adopt a descriptive route guidance strategy within their subnetworks. Information on the least travel time sub-routes within the subnetwork of a controller or a team of controllers at the different time intervals is assumed available. A vehicle that enters this subnetwork is assumed to fully comply with the recommended sub-route, if different from its initial sub-route. Let \( r^\tau_{ijk} \) denote the number of vehicles using sub-route \( k \) departing interval \( \tau \) between origin-destination pair \( i \rightarrow j \) as their initial (historical) route. The number of vehicles that uses this sub-route after applying the route guidance control strategy is denoted by \( q^\tau_{ijk} \). Each local controller or team of controllers is assumed to be equipped with a dynamic traffic assignment (DTA) capability to determine the routing strategy within their boundaries. This DTA capability requires information on the time-dependent demand distribution in the subnetwork of each controller or team of controllers. As mentioned above, each local controller is assumed to be able to estimate the traffic demand within its boundaries. Several solution methodologies are proposed in the literature for the dynamic OD demand estimation problem (Ashok & Ben-Akiva 2000; Bierlaire & Crittin 2004; Cascetta et al. 1993; Ding et al. 1997; Dixon & Rilett 2000; Kang 1999; Nie & Zhang 2008; Sherali et al. 1997; Sherali & Park 2001; Tavana 2001; Tavana & Mahmassani 2001; Xu & Chan 1993; Zhou & Mahmassani 2007). For example, the pioneer work of Cascetta et al. (1993) provides a generalized framework for the dynamic OD demand estimation problem. The framework provides a least squares error formulation for estimating the demand in a general network by making use of the modeling results in the field of DTA. The notation of link-flow proportions is introduced to describe the fraction of OD flow that contributes to the flow on a link in every observation time interval. This approach is used to estimate the local demand in the subnetwork of each local controller \( c \in C \) using the real-time traffic observations \( y_{at} \) and historical assignment matrix \( p_{at}^{c-i|j} \). The demand estimator \( \hat{r}_{ij}^{\tau} \) could be obtained by solving the non-linear mathematical program given below (Cascetta et al. 1993). It minimizes the square of the difference between the observed traffic count and that obtained from the estimated demand for all links and observation intervals.

\[
\text{Minimize} \quad \sum_{a} \sum_{t} (c_{at} - \sum_{\tau} \sum_{f} \sum_{j} p_{at}^{c-j|\tau} \hat{r}_{ij}^{\tau})^2 \\
\text{Subject to:} \quad \hat{r}_{ij}^{\tau} > 0 \quad \forall i, j, \tau
\]  

(3) (4)

If a team of controllers is formed, local controllers in this team share information on their traffic demand. The lead controller of the team uses this information to estimate the demand for the team’s subnetwork (Etemadnia & Abdelghany 2009, 2010). Following the approach
described in (Etemadnia & Abdelghany 2010), a hierarchal recursive mechanism is proposed where the demand of each two adjacent subareas is merged. The merging process continues recursively until the demand for the entire area is obtained. It assumes that traffic crossing from one subarea to another is distributed among the different destinations in the adjacent subarea according to the demand distribution pattern in this adjacent subarea. For example, if two adjacent subareas \( c \) and \( c' \) are connected through node \( e \) (i.e., a boundary node for both subareas), and origin node \( i \in N^c \) and destination \( j \in N^{c'} \), the demand from \( i \) to \( j \) with respect to the merged area \( m \) is estimated using the equality relationship in (5). The index \( \tau \) is the departure time interval at which the traffic departing from origin \( i \in N^c \) arrives at the boundary node \( e \).

\[
P_{ij}^m = \frac{N_c \tau_{ij}^c \cdot \hat{N}_e^c_{ij}}{\sum_{j} \tau^{c'}_{e_j} \cdot \hat{N}_e^{c'}_{e_j}} \quad \forall i, j, \tau, i \in N^c, j \in N^{c'} \quad (5)
\]

The term \( \hat{N}_e^c_{ij} \) is the demand in the subarea \( c \) from node \( i \in N^c \) to the boundary node \( e \). This demand is assumed to be distributed across all destination nodes \( j \in N^{c'} \) following the distribution pattern in \( c' \). The distribution pattern is captures by the ratios \( \frac{\tau^{c'}_{e_j}}{\sum_{j} \tau^{c'}_{e_j}} \cdot \hat{N}_e^{c'}_{e_j} \) \( \forall j \). Once the demand of one controller or a team of controllers is estimated, historical information on the time-dependent demand pattern can be used to predict the demand for the horizon under consideration \( h \). In most practical applications, this time-dependent pattern is approximated by the time-dependent observed traffic counts of some representative links in the network. For example, representing the time-dependent pattern in the form of an autoregressive model of order \( P \), the demand of the future horizon \( d_h \) is estimated using that of the predecessor ones \( d_{h-p} \) using the following relationship:

\[
d_h = \beta + \sum_{p=1}^{P} \phi_p d_{h-p} + \varepsilon_h \quad (6)
\]

Where, \( \beta \): is a constant, \( \phi_1, \ldots, \phi_P \) are the estimated parameters and \( \varepsilon_h \) is a random error term.

Determining the demand distribution for the horizon of interest, the DTA capabilities of each controller/team of controllers is activated to determine the routing strategy that can be adopted considering the prevailing traffic conditions. These prevailing conditions could be significantly different from those based on which an initial route is selected by a traveler. For instance, if one or more incidents occurred in the network, the historical routes for considerable number of the travelers could be subjected to excessive delays. The architecture allows using different routing strategies that could be either descriptive or normative. For example, a descriptive-based routing strategy could reroute the traffic to follow the least travel time routes at the time it enters the subnetwork. Alternatively, the dynamic user-equilibrium (DUE) traffic pattern could be used to provide a descriptive route guidance strategy. The solution of the dynamic system-optimal (SO) traffic assignment problem, including its constrained versions, could be used to derive the normative-based strategies (Peeta & Mahmassani 1995a). Of course, the success of any adopted routing strategy depends primarily on the vehicles' level of compliance to the provided routing instructions.

**COALITION STRUCTURING**
The team forming problem aims at determining the optimal organization of the local controllers in teams assuming that these controllers are adopting a predefined control strategy to improve the performance of the entire network. Define $\mathcal{C}$ as a set of local controllers that are distributed in the network. We also define $\hat{\mathcal{C}}$ as the set of teams that can be configured at any time. We use the binary variable $y_{c\hat{c}}$ to describe the distribution of local controllers among teams. The variable $y_{c\hat{c}}$ is equal to one if controller $c \in \mathcal{C}$ is part of team $\hat{c} \in \hat{\mathcal{C}}$ and zero otherwise. We use $Y_{\hat{c}}$ to describe the set of controllers that are part of team $\hat{c} \in \hat{\mathcal{C}}$. If controller $c$ joins team $\hat{c}$, its contribution in this team is represented by $w_{c\hat{c}}(Y_{\hat{c}})$. The contribution $w_{c\hat{c}}(Y_{\hat{c}})$ is measured as the improvement in the overall network performance as controller $c$ joins team $\hat{c}$ given that $\hat{c}$ consists of the set of controllers $Y_{\hat{c}}$. An optimal team configuration will assign local controllers to teams such that the improvement in the overall network performance is maximized. The mathematical program given below is used to formally define this problem.

Maximize

$$\sum_{c} \sum_{\hat{c}} w_{c\hat{c}}(Y_{\hat{c}}) \ y_{c\hat{c}}$$ \hspace{1cm} (7)

Subject to:

$$\sum_{\hat{c}} y_{c\hat{c}} = 1 \hspace{1cm} \forall \ c \in \mathcal{C}$$ \hspace{1cm} (8)

$$y_{c\hat{c}} \in \{0,1\} \hspace{1cm} \forall \ c \in \mathcal{C}, \hat{c} \in \hat{\mathcal{C}}$$ \hspace{1cm} (9)

The objective function in (7) describes the overall system performance associated with clustering the local controllers in teams. It maximizes the improvement associated with including an autonomic controller $c \in \mathcal{C}$ in a team of controllers, $\hat{c} \in \hat{\mathcal{C}}$. This improvement is expected to be a function of other autonomic controllers $Y_{\hat{c}}$ that join that team and their roles in the team. Constraints in (8) guarantee that each controller is part of one team at any time instance. Decision variables $y_{c\hat{c}}$ are binary integers variables as shown in (9). The optimal coalition structure problem is proved to be a NP-hard problem as the number of possible coalitions grows exponentially with the increase in the number of controllers in the network (Garey & Johnson 1979). Several solution approaches are proposed for this problem. One approach that is used in the literature to study the team formation problem is game theory (Sandholm & Lesser 1997). Game theory is used to describe which coalitions can form in a game of several agents, and how the agents will distribute the benefits of collaboration among themselves. However, game theory does not provide algorithms that agents can use to form the coalitions. Instead, given a set of coalitions, it is used to verify the stability of these collations and check the fairness of benefits distribution (Shehory & Kraus 1995). The coalition structuring problem is also modeled in the form of a Set Partitioning Problem (SPP). The problem involves assigning a group of agents to teams such that each agent falls in one team. However, the SPP is a NP-hard problem (i.e., number of coalitions that can be formed using a agents is $2^a$) (Garey & Johnson 1979). Thus, as the number of agents increases, the time required to determine the optimal structuring grows exponentially. In addition, heuristic approaches that are used to solve the problem usually limit the number of predefined subgroups, which could impact the quality of the generated solution. Zakarian and Kusiak (1999) presented a structure for forming a multi-functional team in a multi-
task environment. The goal is to develop optimal teams such that each team combines the skill set required for that team. They developed a mathematical model to verify optimal composition of multiple teams such that each team satisfies a pre-defined set of requirements (i.e., the team holds a set of skills). They adopted a solution approach that integrates the Analytical Hierarchy Process (AHP) and the Quality Function Deployment method (QFD). This hierarchical methodology permits a large degree of freedom in the formation of teams, where additional tasks can be easily added to the team selection hierarchy without reconstructing the entire model. Gaston and desJardins (2005) proposed strategies for agent-organized networks, and evaluate their effectiveness for increasing organizational performance. The research addresses two main questions: a) how do local agents perceive the collective performance of the entire system? and b) what are best strategies for each local agent to join or leave a team? In this work the authors illustrated the importance of understanding the interaction among agents in order to be able to develop efficient teams.

In this research work, we adopt an algorithm that is based on the work of Shehory and Karus (1995). The algorithm reduces the complexity of the problem by setting an upper bound (k) on the maximum number of team members that can join any team. Figure 3 describes the main steps of this algorithm as adopted in the autonomic traffic management architecture. The algorithm starts by defining potential team leaders in the system. In the current implementation, controllers are assumed to be equipped with incident detection capabilities. A controller that detects an incident is elected to lead a team of controllers to manage the traffic in the vicinity of this incident. Each team leader is capable of configuring a team of neighboring controllers. A pre-defined communication protocol among controllers is assumed. This communication protocol specifies all necessary communication messages to facilitate information sharing and collaboration among controllers. For example, for the purpose of team formation, defined messages in this protocol include invitation-to-join-team, accept-team-invitation and reject-team
invitation. The invitation-to-join-team message is sent from a team leader to neighboring controllers to join its team. The accept-team-invitation or reject-team-invitation messages is sent as a response from a controller to indicate if it accepts or rejects this invitation. Team leaders evaluate the benefits associated with each possible team they could form (i.e., coalition value). The coalition value is computed as the system-wide benefits resulting from collaboration among the team members minus any communication overhead (cost) associated with coordinating the actions of these members. This coalition value is measured in terms of the saving in the network travel time associated with adopting the traffic management plan that is recommended by the team. Without loss of generality, we ignore the communication overhead among controllers in the current version of the algorithm. Once the coalition values associated with all potential teams are evaluated, each team leader announces the best coalition value among all its formed coalitions. The coalition with the best value among all leaders is recommended for activation. All controllers that are part of this coalition are eliminated from the list of controllers. Remaining leaders repeat the process by configuring their potential teams using the set of unused controllers. The process continues until no coalition with a positive value can be formed, or all controllers are allocated to teams.

EXPERIMENTS AND ANALYSIS

In order to examine the performance of the developed architecture, a set of offline experiments are conducted using a roadway network that represents a part of the Dallas metroplex. As illustrated in Figure 4, the study area includes a section of Dallas North Tollway (DNT) and a set of parallel and perpendicular arterials. The network consists of 365 links and 128 intersections/junctions. A demand pattern that approximately represents a typical morning rush period is considered. Following this pattern, the majority of the simulated traffic (≈90%) is assumed to move from the north boundaries to the south boundaries of the area (towards Dallas' downtown). Three congestion levels are considered: low, medium and high with a total loading rate of about 10,000, 14,000 and 18,000 vehicles per hour, respectively. The demand is assumed to be loaded over one hour period and to follow a symmetric triangular loading pattern. DNT represents the principal arterial in the network and is assumed to carry about two-thirds of the traffic in the southbound direction. The other one-third of the southbound traffic is assumed to use all parallel arterials. This traffic split along all southbound arterials matches the observed traffic volumes along these corridors. Several non-recurrent congestion scenarios are considered.

In these scenarios, three roadway incidents are assumed to occur in the network. As illustrated in the figure, the first incident is assumed to occur along the southbound section of DNT between Walnut Hill Lane and Northwest Highway and blocks two of the three available lanes at this section. The second and the third incidents occur along the southbound sections of Hillcrest road and Preston road, respectively. The second incident blocks one lane out of the two available lanes, while the third incident blocks two of the three available lanes. All incidents are assumed to last for 35 minutes, and to start 20 minutes after the start of the demand loading period. As described above, controllers could be integrated in teams to develop efficient traffic management schemes to recover the detected incidents. The used coalition structuring algorithm sets up an upper bound on the maximum number of controllers that could join one team. Different team sizes are expected to result in different traffic management plans with different levels of success in alleviating congestion associated with the incidents. In these experiments, we consider generating traffic management plans based on the collaboration of teams with different sizes. As illustrated in Figure 4, for the incident on DNT, three maximum team sizes are considered: large, medium and small. The upper bounds of these team sizes are determined by the number of controllers that are located within the specified areas. We assume that a controller
manages the traffic in one block in the managed area as illustrated in Figure 4. Thus, for the large team size case, a maximum of seven controllers could be used to form the teams. In the medium and small cases, the maximum number of controllers is set to four and two, respectively. All teams are assumed to provide travelers in their areas with a descriptive route guidance strategy in terms of routes with the least travel times. All vehicles are assumed to be equipped to receive route guidance instructions and all drivers are assumed to follow a bounded-rational route choice behavior. In other words, if a driver is assumed to comply with the information, this driver switches to a new route only if the travel time of this new route is less than that of the historical route by a given threshold (10% saving). The experiments examine the effectiveness of the management plans considering four different drivers' compliance rates with the provided information. These rates are %0 (i.e., all drivers follow their original historical routes), %25, %50, 75%, and %100 (i.e., all drivers comply with the provided route guidance instructions). As illustrated in Figure 4, three screen lines are defined in the study area. These screen lines allow tracking the drivers’ diversion decisions as function of the route guidance instructions that are provided by the different controller teams. The results presented in this section aim at illustrating the performance of the coalition structuring algorithm considering different factors related to the network congestion levels, the spatial distribution of the incidents in the network, drivers’ compliance rates and the upper bound of the controller team sizes. Results related to the demand estimation for the teams can be found elsewhere (Etemadnia & Abdelghany 2009, 2010).
In all these experiments, a traffic-simulation assignment model is used to evaluate the network performance considering the routing scheme that is generated by the different teams (Abdelghany and Mahmassani 2001). The model simulates an urban transportation network at a high level of detail. It consists of several modules including (a) demand generation, (b) travel behavior, (c) shortest path algorithm, (d) vehicle simulation, and (e) statistics collection. The model can accept as demand input either a file listing the population of travelers; their attributes (including origin, destination, time of departure); or a pre-specified, time-dependent origin-destination trip table. Each generated traveler is assigned a set of attributes, which includes his/her trip starting time, generation link, final destination, and a distinct identification number. Prevailing travel times on each link are estimated using the vehicle simulation component, which moves vehicles following the average prevailing speed. The estimated travel time is in a route decision module that is activated at fixed intervals to provide travelers with a set of route options. The activation interval (usually in the range of three to five minutes) is set so that the variation in network conditions is captured, while retaining desirable computational performance. Drivers evaluate the different route options and choose the preferred one. Vehicles are then moved in the network subject to the prevailing traffic conditions until they reach their final destinations along the pre-specified routes. If a driver receives en-route information and this driver is assumed to comply with the provided information, the route of this vehicle is updated accordingly. The model produces different statistics to evaluate the network performance.

Table 1 presents the results of the first set of experiments. It examines the effect of the maximum number of controllers that can be included in any formed team (maximum team size) on the performance of the most efficient team that is obtained through the algorithm's search process. In this set of experiments, the three different maximum team sizes described above are considered (cases a, b and c in Table 1). The incident on DNT is assumed to occur in the network. In addition, a medium congestion scenario is considered and all drivers are assumed to fully comply with the provided information. The performance (coalition value) of all formed teams are compared in terms of the saving in the average vehicle travel time. These performances are evaluated using the traffic network simulation model described above. The table shows the best formed team for each considered maximum team size, and the percentage saving in the average travel time compared to do-nothing scenario. In addition, it gives the number of evaluated teams in each case and the associated running times till the most optimal team is determined. As given in Table 1, the numbers of generated teams are 14, 37 and 597 for the small, medium, and large areas, respectively. One can notice the exponential increase in the number of teams as the upper bound of the number of controllers increases. This exponential increase in the number of teams is also reflected in the recorded running times. It is worth mentioning the running time is mainly governed by the running time of the DTA-based network simulation model that is used to accurately evaluate the effect of the route guidance scheme that is generated by each team. As illustrated in the figure, for cases (a) and (b) which include limited number of controllers, the most efficient team includes all controllers in the specified area. For case (c), some of the controllers are excluded from the most efficient team as the inclusion of these controllers resulted in less efficient routing for vehicles impacted by the incident. In addition, increasing the upper bound of the team size is shown to result in a more efficient team with larger travel time savings. A percentage saving in the average travel time of about 8% is recorded in case (a). This saving increases to about 9.2% and 11.8% in cases (b) and (c), respectively, as the upper bound on the maximum number of controllers increases. The results also show the recorded traffic volumes at the three screen lines. In case (a), the team of
controllers diverted a portion of the traffic from DNT to the frontage road. In cases (b) and (c), the controllers in the team extend to cover Preston road which is also a good alternative route to the blocked section of the freeway. Providing route guidance instructions that include Preston road in addition to the frontage road provided better distribution of the traffic and increased the saving in the average travel time. Comparing cases (b) and (c), the team in case (c) does not seem to divert traffic to Inwood road. The addition saving in case (c) is mainly due to diverting a slight portion of the traffic to the minor road between the frontage road and Inwood road.

Table 1: The Effect of Different Coalition Sizes

<table>
<thead>
<tr>
<th>Re-routing Modules:</th>
<th>Re-routing Modules:</th>
<th>Re-routing Modules:</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of travelers in north-south</td>
<td>% of travelers in north-south</td>
<td>% of travelers in north-south</td>
</tr>
<tr>
<td>north-south</td>
<td>north-south</td>
<td>north-south</td>
</tr>
<tr>
<td>Inwood</td>
<td>DNT</td>
<td>Preston</td>
</tr>
</tbody>
</table>

*ASTT= %8, **RT=35.1, ***NET=14(8,14)  
*ASTT= %9.2, RT=94.6, NET=37(8,18)  
*ASTT= %11.8, RT=1691, NET=597(8,27)  

*ASTT : Percentage of Average Travel Time Saving  
**RT : Running Time (minutes)  
***NET : Number of Evaluated Teams (Min members, Max members)

Table 2 illustrates examples of some of the teams that are evaluated during the search process to determine the most efficient team. The results considers the case with the largest maximum team size. The table shows the area (in gray) that each team of controllers is covering. The average travel time associated with activating each of these teams is also given. As shown in the table, forming teams using the incorrect set of controllers could result in low network performance as these controllers do not provide the suitable alternative routes to manage the traffic during the incident. For example, an average travel time of 35.27 minutes is recorded when Team A is configured. The alternative routes provided by these controllers do not
efficiently accommodate the diverted traffic due to the incident. Team F is the most efficient team for this incident scenario with an average travel time of 29.37 minutes.

Table 2: Example of Generated Team Configurations during the Search Process

<table>
<thead>
<tr>
<th>Team</th>
<th>ATT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team A</td>
<td>ATT: 35.27 minutes</td>
</tr>
<tr>
<td>Team B</td>
<td>ATT: 34.32 minutes</td>
</tr>
<tr>
<td>Team C</td>
<td>ATT: 34.16 minutes</td>
</tr>
<tr>
<td>Team D</td>
<td>ATT: 33.83 minutes</td>
</tr>
<tr>
<td>Team E</td>
<td>ATT: 32.45 minutes</td>
</tr>
<tr>
<td>Team F</td>
<td>ATT: 29.37 minutes</td>
</tr>
</tbody>
</table>

One can expect the network performance to depend on the assumption of how drivers are complying with the provided route guidance instructions that are generated by the different teams. Table 3 gives the results of a set of experiments in which the network performance is recorded considering different drivers' compliance rates. In this set of experiments, we consider a scenario in which an incident is occurring along DNT with a medium congestion level. The upper bound on the maximum team size that can be formed is set to medium. Drivers' compliance rates with the provided route guidance instructions are set to change from 0% to 100%. The results show the percentage saving in the average travel time compared to that of the case of 0% compliance rate. In addition, the structure of the most efficient team for each case is also presented. For all cases, the traffic volumes along the screen line downstream the incident are recorded. As shown in the table, the highest percentage saving in the average travel time (10.40%) is recorded at a compliance rate of 50%. In this scenario, about 12.5% and 18.3% of the drivers are estimated to use the frontage road and Preston road (the two alternative routes at the incident location). The case in which 25% of the drivers are assumed to comply with the information gives a saving in the average travel time of 6.04% implying that not enough traffic is
diverted to the recommended alternative routes. As shown in the table, the percentage of vehicles using the frontage road and Preston road in this case are recorded to be 4.5% and 15.6%, respectively. In the scenario with 100% compliance rate, the percentages of drivers that use these two alternative routes increase to 7.6% and 20.5% implying that a situation of overreaction to the provided information has occurred. The overreaction to the provided information is associated with a reduction in travel time saving to 9.25%.

Table 3: Network Performance under Different Compliance Rate with the Provided En-Route Information

<table>
<thead>
<tr>
<th>CR</th>
<th>ATT</th>
<th>ASTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>33.3</td>
<td>0%</td>
</tr>
<tr>
<td>25%</td>
<td>31.29</td>
<td>6.04%</td>
</tr>
<tr>
<td>50%</td>
<td>29.84</td>
<td>10.4%</td>
</tr>
<tr>
<td>75%</td>
<td>29.86</td>
<td>10.3%</td>
</tr>
<tr>
<td>100%</td>
<td>30.22</td>
<td>9.25%</td>
</tr>
</tbody>
</table>

*CR: Compliance Rate to provided en-route information  
**ATT: Average Travel Time (Seconds)  
***ASTT: Percentage of Average Travel Time Saving

The next set of experiments examines how the change in the network congestion level could impact the configuration of the optimal team and the associated network performance. In these experiments, the incident along DNT is considered and the maximum number of controllers in the team is set at the medium level. Drivers are assumed to fully comply with the provided en-route information. The three different congestion levels described above are considered. The results of this set of experiments are given in Table 4, which present the
configuration of the optimal team and the associated percentage saving in the travel time. As shown in the table, the result indicates that the configuration of the optimal teams is a function of the network congestion level. Furthermore, the highest percentage saving in the average travel time is recorded for the medium congestion case (9.2%). For the high congestion scenario, the saving in the average travel time was limited to less than 2%. This can be explained by the fact that all parallel southbound routes are highly congested and hence no better rerouting options are available. A surprising slight increase in the average travel time is recorded in the low demand case. This increase could be contributed to the nature of the provided route guidance information as it is generated based on the current travel time when the vehicle enters the team's subnetwork rather than on the corresponding anticipated travel time. Such information could unnecessary reroute few vehicles to longer routes.

Table 4: Network Performance under Different Demand levels

<table>
<thead>
<tr>
<th></th>
<th>High Demand</th>
<th>Medium Demand</th>
<th>Low Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inwood</td>
<td>ScrLine_2</td>
<td>ScrLine_1</td>
</tr>
<tr>
<td></td>
<td>DNT</td>
<td>Preston</td>
<td>Hillcrest</td>
</tr>
</tbody>
</table>

The last set of experiment presents the network performance considering scenarios in which multiple incidents are assumed to occur simultaneously in the network. Two scenarios are considered. In the first scenario, two concurrent incidents that are spatially independent are assumed. These two incidents are located along the south bound of DNT (incident 1) and the southbound of Hillcrest road (incident 2). In the second scenario, the two incidents are assumed to be in the proximity of each other. The first incident is located along the south bound of DNT (incident 1) and the southbound of Preston road (incident 3). Tables 5 and 6 illustrate the locations of these incidents and the area of the maximum team size for each incident. The capacity reduction and the durations of these incidents are as described above. In both scenarios, a medium demand level is assumed. In addition, four information compliance rates of 25%, 50%, 75% and 100% are considered. Table 5 gives the results of the first scenario. The percentage in the average travel time saving is given for the different information compliance rates. The travel time saving is given when only a team is formed for the incident on DNT, and when two teams are formed for the two considered incidents. As shown in the table, considering the severity of the DNT incident, activating the team for this incident results in most of the saving in the
average travel time. An additional slight travel time saving is recorded when the team on Hillcrest road is activated. For instance, an average saving of about of 9% is recorded when the team for the DNT incident is activated assuming 100% compliance rate. The saving increased to 9.5% when the team for the Hillcrest road is also activated.

Table 5: Network Performance under Two Spatially Independent Locations of Incidents

<table>
<thead>
<tr>
<th>% Compliance Rate to En-route Information</th>
<th>% Average Travel Time Saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Team for Incident 1</td>
<td>5</td>
</tr>
<tr>
<td>Two Teams for Incidents 1&amp;2</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 6: Network Performance under Two Spatially Inter-dependent Location of Incidents

<table>
<thead>
<tr>
<th>% Compliance Rate to En-route Information</th>
<th>% Average Travel Time Saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Team or Incident 1</td>
<td>5</td>
</tr>
<tr>
<td>Two Teams for Incidents 1&amp;3</td>
<td>7</td>
</tr>
<tr>
<td>Combined Team for Incidents 1&amp;3</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 6 gives the results for the second scenario in which two nearby incidents are assumed to occur simultaneously. The results are given for three different cases related to the configured teams: a) one team is activated for the incident on DNT; b) two independent teams for the two incidents; and c) one large team to simultaneously manage both incidents. As shown in the table, a saving in the average travel time is recorded for all team configurations and all
assumed compliance rates. As for team configurations, the highest network performance is recorded when one large team is configured to manage both incidents. Configuring two independent teams for the two incidents result in the least network performance. This performance is even less than that when one team is configured for the main incident on DNT. At 50% compliance rate, a saving of 8.6% is recorded when one large team is configured for both incidents. For the same compliance rate, configuring two independent teams for the incidents reduces the saving to 6.2%. When only one team is configured for the DNT incident, a saving of 7.9% is recorded. As described above, the coalition structuring algorithm follows a greedy-based approach. The optimal team for DNT is first determined. While this team is determined, the actions of all subsequent teams that might be formed are ignored. As illustrated in the scenario presented in Table 5, if the subsequent teams that might be formed are far from the current team, the effect of the actions of these teams on the first formed team is expected to be minor. On the contrary, considering the results obtained in Table 6, when the subsequent teams are close to the first team, the actions of these teams might affect those of the first formed team and hence affect its performance.

CONCLUSIONS

The paper proposes an autonomic architecture for real-time traffic network management. Following this architecture, controllers are assumed to share information on the observed traffic pattern and their control actions. In addition, controllers could be dynamically configured to operate in teams to develop integrated control strategies that best cope with the observed traffic pattern in the network. The paper presented the results of off-line simulation-based experiments that illustrate the performance of the proposed architecture. Based on the obtained results, more efficient routing could be obtained through collaboration among individual controllers, which result in considerable travel time savings. In addition, increasing the upper bound of the maximum team size generally result in configuring efficient teams with more travel time saving. The structure of the configured teams and their performances are also shown to be a function of the network congestion and the drivers’ compliance rate with the provided information. The results also show that efficient team configurations can be obtained when multiple incidents occur in the network. If the incidents are close to each other, the results suggest forming one large team that manages both incidents. Several extensions for this research work are underway. For example, research work is ongoing to provide the operational models for the system's local controllers including their semantic and communication protocols. In addition, in the current research work, controllers are assumed to adopt a simple descriptive re-routing strategy. The use of more elaborate (normative) re-routing strategies is considered as an extension of this research work. Finally, work is undergoing to implement and test the real-time version of the proposed architecture.

REFERENCES


