Geospatial Assessment of Terrestrial Communication Infrastructure Requirements during the Process of Traffic Signal System Upgrade

Milos Mladenovic¹, Montasir Abbas, PhD²
1. Graduate Research Assistant, Virginia Polytechnic Institute and State University, USA, Phone: (540) 553-5949, E-mail: milosm@vt.edu
2. Associate Professor, Virginia Polytechnic Institute and State University, USA

Abstract: Departments of Transportation (DOTs) resort to upgrading traffic signal control infrastructure in response to growing traffic needs and maturing technology. These upgrades are infeasible to conduct instantaneously, requiring the development of a spatial and temporal migration plan. The upgrade of signal control equipment could potentially change the requirements of the supporting communication infrastructure. The essential need for analysis of communication infrastructure is related to the idea of effective resource planning and improved decision-making process, along with reduction of investments and work duplication among DOT staff. This research focuses on the development of framework and methodology for such an analysis during the creation of a migration plan for the upgrade of signal control system. The developed tool integrates Geographical Information Systems techniques with telecommunication procedures for quality of service analysis.

1 Introduction

Communication infrastructure is considered as an integrated part of any modern Intelligent Transportation System (ITS) [1]. This communication infrastructure defines logical and physical connections intended to fulfil the communications requirements of a given transportation agency [2]. On the other side, traffic signal control subsystem is an essential and very important constituent of any agency’s ITS. Consequently, the communication infrastructure is a highly important part of a particular traffic signal control subsystem.

1.1 Migration Plan for Signal Control Equipment

Traffic signal control infrastructure can be a too large system for instantaneous upgrade. Because of that complexity, the analysis of this process should undertake an established systems-engineering approach conducted in several parts [3]. For creating an optimal migration plan, this process would initially require identification of gaps in the system features. In addition, the process would require the evaluation of future control equipment and the development of a stepwise migration plan with a long-term replacement perspective. This migration plan would specifically define the points in time and locations for the upgrade of evaluated field equipment. As a part of this high-level process, assessment of communication infrastructure is directly related to geospatial analysis (Figure 1).
1.2 Research focus and approach

Development of an optimal migration plan for such system is not a simple task since a signal control subsystem consists of many elements. From the high-level perspective at migration plan development, requirements of changing traffic demands are the most important issue. This points out to the decision on the upgrade of immediate control system elements (such as traffic signal controllers) as a primary in the migration schedule. After the decision on the upgrade of major control components, there is the need for assessment of new communication requirements. The logic of integrated infrastructure management requires that communication system should expand, if needed, along with new signal-control system installation. The idea behind this logic is the intention for improved system components integration, minimization of investments and resources, and reduction of work duplication among transportation agency staff.

Taking into consideration the communication infrastructure as a physical structure with a wide spatial range and large database, its analysis is recognized as Spatial Data Integration issue [4]. Since communication infrastructure has a strong spatial component, the intuitive cognitive power of a decision maker referenced to this spatial component has to be utilized in the analysis. Temporal and spatial identification of the data used as information in the decision-making actions is recognized as one of the critical steps in the process. In essence, there is a need for the ability to manipulate spatial data in different forms and extract additional meaning from them. This should provide crucial information needed to support the traffic engineers through the process of system upgrade. The set of tools recognized as capable to fit these requirements are known as Geographical Information System (GIS).

2 Synthesis of past knowledge and efforts

2.1 Related Telecommunication Network Features

By the definition, the basic purpose of a telecommunication network is transmission of information from one user/device to another user/device of the network [5]. The
communication infrastructure specific to ITS in North America bases on the National Transportation Communications for ITS Protocol (NTCIP) framework. NTCIP is a group of general-purpose communications protocols and transportation-specific data dictionaries/message sets that support Centre-to-field (C2F) or Centre-to-Centre (C2C) applications [6]. NTCIP has plant, subnetwork, transport, application, and information level. The plant level represents the actual physical infrastructure. The subnetwork level has standards that define procedures for exchanging the data between two devices over some communication media. The transport level standards define procedures for exchanging and routing the data on the network level. The application level standards define the procedures for exchanging data on the level of application. The information level has standards that define the meaning of data and messages exchanged between Traffic Management Centre and field device or another Centre. Each level includes specific protocols related to the either C2F or C2C application.

Modern telecommunication analysis bases on standardized Open Systems Interconnection (OSI) model [7]. This model has seven layers: physical, data link, network, transport, session, presentation, and application. The physical layer coordinates functions required to carry a bit stream over a physical medium. The data link layer transforms the physical layer to a reliable link. The network layer is responsible for source-to-destination delivery of a packet, possibly across multiple networks. The transport layer is responsible for process-to-process delivery of the entire message. The session layer establishes, maintains, and synchronizes the interaction among communicating systems. The presentation layer is concerned with the syntax and semantics of the information exchange. Finally, the application layer enables the user, either human or software, to access the network.

2.2 *Telecommunication infrastructure specific to signal control*

The ITS communication infrastructure bases on a similar layered structure as OSI model. National Transportation Communications for ITS Protocol (NTCIP) framework (Figure 2) defines this infrastructure. NTCIP is a group of general-purpose communications protocols and transportation-specific data dictionaries/message sets that support Centre-to-field (C2F) or Centre-to-Centre (C2C) applications [6]. NTCIP guides and defines most of the ITS telecommunication activities. NTCIP has plant, subnetwork, transport, application, and information level. The plant level represents the actual physical infrastructure. The subnetwork level has standards that define procedures for exchanging the data between two devices over some communication media. The transport level standards define procedures for exchanging and routing the data on the network level. The application level standards define the procedures for exchanging data on the level of application. The information level has standards that define the meaning of data and messages exchanged between Traffic Management Centre and field device or another Centre. Each of the levels includes specific protocols related to the either C2F or C2C application and consequently those protocols determine the throughput requirements introducing overhead bits.

2.3 *Service capability in telecommunication networks*

Elements and protocols of the previously mentioned NTCIP layers directly affect the communication throughput. Throughput, or data rate, is the number of data elements (bits)
sent in 1 second and is a measure of how much information can be sent through a network connection. This data-carrying capacity is expressed in bits per second (bps). Theoretically, throughput is in direct relation to the bandwidth. Bandwidth is the range of frequencies used for transmission and is the major factor that determines the carrying capacity of a telecommunications channel. In practical applications, the term bandwidth is often interchangeable with throughput due to their close relationship. In addition to the available bandwidth, throughput depends on the level of signals used, and the quality of a channel (level of noise). The basic equation of Shannon’s capacity (1) that represents a throughput of a network element is defined as:

\[
C = \frac{T}{2} \cdot log_2(1 + \frac{S}{N})
\]  

(1)

where:
- \(C\) – capacity,
- \(T\) – symbols in time,
- \(S\) – signal power,
- \(N\) – noise power

Throughput of the network is the number of packets passing through the network in a unit of time. Beside throughput, there are other quality-of-service parameters involved in the analysis. These parameters are mainly related to communication redundancy and reliability. Redundancy relates to error detection and correction. These processes are usually conducted through the introduction of additional information into transmitted data code, as defined by the protocols on different OSI layers. In addition, reliability relates to network ability to provide uninterruptable service. This issue is usually addressed with automatic traffic redirection, using algorithms that operate in dynamic load conditions. The routing of data packets through the network is the responsibility of the network OSI layer (ex. Internet Protocol).

2.4 GIS applications in transportation and telecommunications

As stated in the research approach section of this paper, GIS is well adapted to respond to particular requests of this research issue. GIS integrates hardware, software, and data for capturing, managing, analysing, and displaying all forms of geographically referenced information [8]. Visualization, data management and geographical modelling capabilities are the main attractions for using GIS for traffic signal control infrastructure management [9].

GIS has applications in the area of transportation at different levels, such as planning, preliminary design, construction, and operations [9-15]. GIS applications in the area of telecommunication started with applications for automated mapping and facilities management [16]. Further research introduced the GIS ability to analyse the adequacy of a regional broadband infrastructure by spatially relating infrastructure facilities with demand (in this case, schools running distance education applications) [17]. Some of the most recent research has implemented GIS in WiMax wireless network planning and performance evaluation [18]. However, there is an evident lack of any significant research in the area of
3 Framework and methodology

In the development of the framework and methodology, an infrastructure assessment links infrastructure facility information with application requirement information in order to reveal a spatial location of infrastructure supply shortage. Furthermore, this framework directly relates and extends to the framework for the migration plan of the signal control equipment. Having in mind that a migration plan has temporal and spatial component, the assessment tool for telecommunication infrastructure has to incorporate similar logic [3]. The integration of two analytical processes is the general course of action. The one part of the process is the calculation of throughput requirements, based on communication analysis. The other part of the process is the spatial data manipulation and visual assessment of network expansion locations, using all the powerful capabilities that GIS can provide. As the basis for this tool, the general course of assessment actions is guided by two important questions in this assessment process:

1. What are the network expansion requirements levels?
2. Where are the locations of network expansion requirements?

The analysis emphasize on the locations where the signal control equipment is first upgraded to the next generation. This research assumes that the optimal solution for signal control equipment exists before the communication infrastructure assessment. It is assumed that spatial and temporal plan for upgrade of signal control equipment is conducted using some of the geospatial techniques, such as GIS. Following this logic, the existing geo-referenced database containing information on signalized intersections can be used in the layered structure as a communication points’ layer. Furthermore, it is assumed that the upgrade of the signal control equipment is done on the zone basis, grouping a definite number of signalized intersections in the zones by their upgrade priority. The final representation of the migration plan for signal control equipment would represent zones with different upgrade priority level. The actual process of defining the zones and calculating the solution on the larger scale is not a part of this research.

The additional layer added before the start of the communication infrastructure assessment is a polyline layer. This layer represents the actual location of communication infrastructure in referenced space. This layer, along with data and decisions from the migration plan is used for overall spatial telecommunication infrastructure analysis. Assumption is that there is available feature data set, which is geo-referenced and ready for analysis having complete spatial representation of communication locations and infrastructure. The spatial, geometric and attribute compatibility of objects used in GIS analysis is disregarded in the methodology development. This includes a large list of incompatibilities, including differences in the measurement techniques, coordinate systems, spatial scale, aerial coverage, data models, data structures, entity sets, attribute measurement scales, as well as temporal coverage. Fig. 2 presents the overall framework flow of analysis.
3.1 Telecommunication procedure for network throughput calculation

In the first step of the throughput calculation, engineers involved in the process must be aware of all of the devices deployed in the system and their communication requirements. These factors are used in the calculation to determine the amount of required throughput. For the actual process of throughput analysis, it is important to gather and consider the relationship among the following key variables:
- Transmission bit rate;
- Transmission method;
- Transmission latency;
- Response delay of the field device;
- Application message size;
- Data element format;
- Frequency of communication; and
- Number of devices sharing the same line or channel;

Data rate analysis bases on previous key variables and has the following steps:
1. Estimate messages exchange frequency;
2. Estimate Application, Transport and Subnetwork Protocol size;
3. Estimate timing factors (delays);
4. Estimate number of drops per channel;

The required application data rate is generally computed as:

\[
Data\ rate = \frac{size\ of\ data\ exchange \theta}{(1 - latency)}
\]

Where:
- size of exchange – the number of bits application is sending along with overhead protocol bits;
- latency – the summation of all the timing factors that delay the instantaneous data transmission.
3.2 Assessment steps

The developed procedure is similar to the process of design of communication systems, having the perspective “from the ground up”. Ten procedural steps, grouped based on the side of analysis where they relate, are presented on the Figure 3.

![Diagram showing procedural steps]

**Fig. 3** Analysis procedure related to GIS or Communication

4 Example application

4.1 Description of the hypothetical communication system

The hypothetical system (with similarities to the one provided in the NTCIP guide [6]) assumes a traffic control system with a distributed communication infrastructure. The existing communication infrastructure fulfills the requirements of the present system, operated by type 170 controllers. The control system is about to be upgraded with newer controller type 2070 that will increase communication overhead. This overhead will then result in higher transmission requirements. All of the signalized intersections are connected in the C2F system, controlled by the central computer located in the Traffic Management Centre (TMC). Communication data transfer is divided between TMC computer and master controllers, and between master and local controllers. This communication mainly accomplishes using Application layer services that convey the requests to access and modify values of master and local controller objects. The actual message consists of a set of data elements or objects and is determined by a specific application layer protocol. The routable connection between the TMC and master controllers, has following NTCIP-based structure:

- Application level - SNMP protocol (without using Traps);
- Transport level – UDP/IP;
- Subnetwork level – Point-to-point;
- Plant level – two pairs per channel dial-up telecommunication line (V-Series modems).
Non-routable connection is used for sending dynamic objects from master to local controllers. The partial NTCIP-based layered structure is:

- Application level – STMP protocol;
- Transport level – T2/NULL;
- Subnetwork level – Point-to-multipoint;
- Plant level – two pairs per channel twisted pair (FSK modems).

In this case, the critical operation is communication between modems in master (primary device) and local controllers (secondary device). This type of operation is multicasting (one-to-many) polling. The primary device initiates a session and determines the use of a communication channel at any given time. The master controller sends a “time hack” to controllers and polls for available data frame from all field units connected to the communications drop. The master controls the locals through broadcast of different data frame (if local controller is not in Free or Flash). The local controller transmits status, database, and system detector information, and receives system commands and data transmission. The assumed system operation is as following:

- System commands that local receives on once per minute basis are:
  - Coordination pattern,
  - Command for Coordinated, Free, standby or Flash mode
  - Time and date,
  - Request for local status,
  - Special function command (added in the upgraded system)

- Local controller is polled by commands once every second with a time hack and query for database/status data:
  - Cumulative volume and occupancy for each detector in the last sample period,
  - Average speed for each speed detector in the last sample period,
  - Green and yellow status for all phases and overlaps,
  - Walk and pedestrian clearance status for all phases,
  - Vehicle and pedestrian detector status,
  - Phase termination status,
  - Local time,
  - Coordination status,
  - Conflict flash status,
  - Local flash status,
  - Automatic flash status,
  - Local Free,
  - Pre-empt activity and calls,
  - Status of user-defined alarms;

- Controller stores data until polled and must response within 20 ms;
- System operates in overlap full duplex mode;
The data circuit-terminating equipment in the non-routable part of network is analog modem, connected to controller over RS232 cable. The modem converts controller’s digital output to analog format voice frequencies. As required by modem specifications, the port may be configured as Half or Full Duplex asynchronous, over a physical switch for controller port 3B. Each controller has assigned unique IP address according to Transmission Control Protocol/Internet Protocol (TCP/IP) standard “dot notation”. This IP address is used to identify each controller in the communication system. The address must be assigned from the same subnet as the other network devices. If the controller is connected to an IP router, the address must be valid for that router. In addition, the controller might have the ability to choose NTCIP standardized or proprietary protocol, but in this example we assume NTCIP is used.

The controller software can support valid bit rates of 1200, 2400, 4800, 9600, 19200, 38400, and 57600 bps, depending on the controller Computer Processing Unit (CPU) and the modem being used [19]. The 2070-6A module can support 1200 bps, and the 2070-6B module can support 9600 bps. Model 400, the standard modem for use with 170 controllers, only supports 1200 bps communications. Modulation used to connect modems in the cabinets to central control computers is Frequency Shift Keying (FSK) for speed below 9600 bps. Phase Shift Keying (PSK) and Quadrature Amplitude Modulation (QAM) are used for higher speeds.

4.2 Spatial analysis combined with telecommunication analysis

Spatial communication infrastructure analysis localizes the area where immediate analysis of communication infrastructure, in the zones identified as having the highest priority for migration plan upgrade. An example of the spatial migration plan for signal control equipment is presented by Figure 4. The data presented in this figure [20] is illustrating migration plan for signal control equipment on a larger area. In the Figure 4, the priority of upgrade for each zone is assigned as number from one to five, with zones having one as those having the highest priority for upgrade and vice versa. Any further analysis should initially focus on the network parts providing communication service to the signalized intersections with the highest upgrade priority. The area circled in blue is in the “red zone” and is part of the communication infrastructure belonging to one communication drop where further analysis focuses.
The features of communication points and infrastructure are included as attributes of GIS layers. Attributes assigned to the controller layer features consist of X coordinate, Y coordinate, Signal_Number, Zone_ID, Upgrade_Priority, Min_Throughput_Existing and Min_Throughput_New. The X coordinate and Y coordinate are related to GPS coordinates of the intersection controller. The Signal_Number is referring to internal DOT identification number for the traffic signal controller. The Zone_ID and Upgrade_Priority are attributes showing zone identification number for upgrade and upgrade priority level for the particular signalized intersection. The Min_Throughput_Existing and Min_Throughput_New provide information on the throughput required by the existing or introduced signal control equipment. The attributes of communication infrastructure layer are Object_ID, Shape_Length, Media_Type and Existing_Throughput. The Object_ID is a DOT identification number for a telecommunication network segment. The Shape_Length provides additional spatial related information. The Media_Type is providing information related to the transmission medium. The Existing_Throughput is the throughput of a specific communication network segment.

First, the engineer needs to determine the upgrade priorities of communication points, using Upgrade_Priority layer attribute. Points with Upgrade_Priority value of one are the start of analysis. GIS Selection by attribute tool selects the points using a Structured Query Language (SQL) relation. The locations of the communication points are then identified, along with their communication devices. They are represented in GIS as point layer. In the third step, we identify the existing communication requirements, that is 800 bps for the present system features. The future system features result in the communication rate of 900 bps, due to the upgrade to controllers having standard with higher data exchange requirements. Next, we identify the location of existing communication infrastructure, along with the information on telecommunication infrastructure for the selected communication points. All the previous information related to the telecommunication infrastructure features
are used for the calculation of throughput required for each communication point. These values are assigned to attributes Min_Throughput_Existing and Min_Throughput_New of the communication point layer.

Further spatial analysis identifies linear segments related to communication points under investigation. These segments are defined and grouped by the location of routing/switching points, i.e. the smallest network segment that can be separately investigated. A value is assigned to attribute Existing_Throughput and in this example all the communication network segments have transmission capability of 9600 bps. A new layer is created in the next step. This layer is a combination of two separate spatial datasets – communication points and communication infrastructure lines and their attributes. The new layer is a polyline with the summary of attributes values from the point’s layer. In essence, the features of one dataset falling within the spatial extend of another dataset are combined and summed. Attribute values are summarized using the SUM operation, available in ArcMap software. In reality, the coordinates of communication points and the polyline representing communication infrastructure might not overlap. However, the option of selecting points that are closest to the polyline can overcome this issue. This way, all the points are assigned to the closest polyline and their attributes are summarized creating attributes of a new layer. Attributes of different layers and their relations are presented by the following Figure 5. The calculation of difference in existing and required future throughput of network segment is conducted in this layer.

Values of attributes Min_Throughput_New and Existing_Throughput have been used for calculation and creation of joined output layer attributes. After the “Join” operation and removing of unnecessary layer attributes, new layer has attributes Object_ID, Zone_ID, Upgrade_Priority, Sum_Min_Band, Sum_Throughput and Available_Throughput. The Sum_Min_Throughput is the summation of the attribute Min_Throughput values from the points closest to a specific line structure. The Sum_Throughput is the actual throughput required by all the devices that are sending the data to the Centre over that communication
network segment. The Available_Throughput is the difference between Sum_Throughput and Existing_Throughput values for specific linear network segment. The Available_Throughput is used at the end as a representative value showing network element transmission capabilities.

### 4.3 Analysis results and improvement recommendations

Figure 6 is a graph obtained from an inbuilt GIS graphical representation tool, showing the values of Available_Throughput after the final calculation using spatial analysis. As depicted by the graph, the Available_Throughput of network element with Object_ID 14 has a negative value. Consequently, this indicates that this network segment has no enough extra throughput left after the system upgrade.

![Available throughput of communication links](image)

**Fig. 6 Available throughput of communication links**

GIS tool Selection by attribute (Figure 7) is used for the spatial representation of assessed critical network infrastructure. SQL query of Available_Throughput attribute of ≤ 0 bps identifies the information on the critical network segment. In theory, communication link could hypothetically allow no overhead throughput available, thus allowing the value of 0 bps of Available_Throughput. In practice, overhead throughput is needed due to the requirement for round trip communication time reduction, signal attenuation, signal-to-noise ratio, and latency induced by a device to properly format and send a response. These are all the reasons why available overhead throughput less or equal to 0 bps is unacceptable in practical applications.
Finally, the analyst can identify the specific network element that does not have required throughput to support communications after a system upgrade (Figure 8). The linear structure requiring expansion of transmission capabilities is marked on the map in blue. Furthermore, the information on a potential throughput problem and its location leads to the recommendations for improvement. Improvements in communication infrastructure for signal control system could be introduced on any of NTCIP layers. Finding solutions can start with redefining system settings or improvement and replacement of different system components, iteratively inside the layered structure of NTCIP. The recommended initial approach is trying to modify the upper layers and then look for improvement in modifying elements in lower NTCIP layers.

The improvement could be obtained by introducing IP over Ethernet and replacing second-by-second polling architecture with Simple Network Management Protocol (SNMP) based on trap commands. SNMP is the application level protocol, based on manager/agent
relations. Essentially, using SNMP trap commands, field device is asynchronously initiating communication only when it has something to send, thus significantly reducing throughput requirements.

In addition, the change could be using of synchronous instead of asynchronous modems since they do not introduce overhead bits. In addition, a change from FSK to PSK/QAM could increase throughput without increase in signaling frequency. In addition, the replacement of analog with digital modems is also one of the possible solutions. A final-resort solution for this example is the investment into Synchronous Optical Network (SONET) fiber cable for replacement of existing copper cables. Fiber optics throughput capabilities can resolve all issues related to the communication in one ITS and are usually utilized as an ultimate solution. The decisions for improvement should base on the previously developed Communication Master plan and its planned expansion.

5 Conclusion and further research

This research is guided by the need for assessment of communication infrastructure during the upgrade of traffic signal control system. The presented research develops a framework and methodology for analysing telecommunication infrastructure using geospatial tools. The process of infrastructure analysis has twofold basis: requirement analysis of communication system along with implementing GIS as a tool that greatly improves the process of spatial analysis. This tool is specifically intended for implementation in the process of signal-control equipment upgrade. GIS is chosen as the basis for the development due to its ability to integrate data and processes related to spatial entities.

One of the goals of this research is to provide a starting point and guidance for similar implementation cases. The final product is a tool that can be potentially utilized in the process of identifying existing infrastructure gaps and the development of recommendations for a DOT’s Communication Master Plan. Finally, having the role of initial research in this field, this paper opens a field for numerous possibilities. Potential advances and development can be primarily found in the area of telecommunication infrastructure analysis and assessment related to traffic signal control infrastructure. One of the first expansions of this research could go toward the area of wireless communications utilized for traffic signal control systems. In addition to this, there is potential for assessment of effects that the introduction of Vehicle-Infrastructure Integration will have on the future communication infrastructure. This assessment could provide the information for adjustment of Communication Master Plans nationwide to support incoming technologies.

6 Acknowledgement

This work was supported by the Virginia Centre for Transportation Innovation and Research (VCTIR) under the project “Evaluation of Merits and Requirements of Next-Generation Traffic Control System for VDOT's Northern Region Existing Infrastructure”.

7 References

5. Green, J.H., *The Irwin Handbook of Telecommunications* 1997: Irwin