INFORMATION THEORETICAL MODELING OF COMPLEX COMMUNICATION NETWORK USAGE PATTERNS

by

STANLEY CARL THOMPSON

MURAT M. TANIĸ, COMMITTEE CHAIR
GARY J. GRIMES
LAURIE L. JOINER
MURAT TANJU
GREGG L. VAUGHN

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2008
INFORMATION THEORETICAL MODELING OF COMPLEX COMMUNICATION NETWORK USAGE PATTERNS

STANLEY CARL THOMPSON

COMPUTER ENGINEERING

ABSTRACT

Today’s voice and data connectivity typically utilizes a complex system of networks of networks. The development of models for characterizing usage patterns in complex networks would be useful in projecting capacity requirements in growing systems. Simplified models applied to a complex network for timing and sizing algorithms would significantly reduce the amount of computations and storage necessary to produce forecasts of capacity requirements. This composition introduces a network usage pattern model based on decomposition by information transfer and verifies its application in real-life voice-data networks.

Complex systems can be modeled and decomposed into sub-systems by observing the interactions among their elements. A normalized transmission parameter is used in this study as the model for comparing sets of measurement data to model instances. Methods for constructing example instances and the method of comparison are described. Measurement data for five voice message trunk groups and ten data circuits is analyzed using three different instances. Validation is accomplished using the model instances to predict the parameters of combinations of traffic usage and comparing the predictions to calculated parameters of the usage combinations.

Results of modeling usage data for telephony trunks and internet usage data for two types of circuits are described. Time-consistent busy-hour model instances are compared to 24-hour model instances for each case. For one of the Internet circuit types,
a third model instance with a 6-hour busy period is included. Time-consistent busy-hour instances had the lowest valued transmission parameters. The 24-hour instances had the highest valued transmission parameters and the 6-hour busy period instances had values in between. Instances with greater transmission parameters yielded more accurate predictions when combinations of measurement data had non-coincident usage patterns.

Study results support the original hypothesis that development of models for characterizing usage patterns in complex networks would be useful in projecting capacity requirements in growing systems. The normalized transmission parameter is a useful predictor of relative accuracy of a model in predicting effects of combining usage on trunks or circuits where there was a significant difference in model instance parameters and trunks or circuits had dissimilar busy hours or busy periods.
DEDICATION

This work is dedicated to my wife, Alethia, for her encouragement, patience, and support.
ACKNOWLEDGMENTS

The efforts of George Jung, Daniel Baeza, PhD, Jason Sutton, and Ryan Sanderson in providing data sources and support in the area of voice and data trunk architecture and sizing theory are appreciated. I appreciate the support and encouragement of my team members, co-workers, and supervisors at AT&T. Also appreciated are the advice and encouragement of my committee, Professors Gregg Vaughn, PhD, Gary Grimes, PhD, Murat Tanju, PhD, and Laurie Joiner, PhD. A special thank you to Yalcin Tuncer, PhD, for the support and advice he provided in the area of statistical methods. The advice, guidance, support, and inspiration of my committee chair and advisor, Murat Tanik, PhD, are especially appreciated. I thank you, Dr. Tanik, for your patience.
TABLE OF CONTENTS

ABSTRACT ............................................................................................................................... iii

DEDICATION ........................................................................................................................... v

ACKNOWLEDGMENTS .......................................................................................................... vi

LIST OF TABLES ................................................................................................................... x

LIST OF FIGURES ................................................................................................................. xi

CHAPTER

1 INTRODUCTION ..............................................................................................................1
   Problem Statement ..........................................................................................................1
   Importance .......................................................................................................................2
   Hypothesis .......................................................................................................................2
   Literature Review ..........................................................................................................3
   Summary ............................................................................................................................5

2 IMPROVING THE ECONOMIC DESIGN OF DATA AND VOICE NETWORK
   ARCHITECTURES USING INFORMATION THEORY .........................................................6
   Economic Design of Voice Network Architectures .........................................................6
   Current Approach Using Time-Consistent Busy Hour (TCBH) ......................................6
   Potential Improvement Using Time-Consistent Multi-Hour (TCMH) .........................8
   Economic Design of Data Network Architectures ..........................................................9
   Current Approach of Over-Sizing Initially and Rebuild at Exhaust ............................9
   Potential Improvements by Projecting Infrastructure Exhaust Timing
   Using Sizing Models .......................................................................................................10
   Other Approaches ..........................................................................................................12
   Summary ..........................................................................................................................13

3 MODELING TRAFFIC USAGE .........................................................................................14
   Comparing the Activity of the Elements of a Complex System to Simplified
   Models of the System Using Information Contained in the Measures .......................14
   Measurement Data May Contain Structure .................................................................14
# TABLE OF CONTENTS (Continued)

## CHAPTER

| Structure Can Be Modeled                                                                 | 14 |
| Models Can Be Compared to Observed Data Using Entropic Measures                        | 15 |
| Entropic Measures                                                                        | 15 |
| Creation of Simplified Models of Internet Circuit Usage Patterns Using Entropic Measures | 17 |
| Examples of Internet Circuit Usage Measurement Data Contain Structure                   | 17 |
| Creating and Comparing Various Models of Internet Circuit Usage Using Entropic Measures | 17 |
| Development and Validation of Models                                                    | 18 |
| Categories of Models                                                                     | 18 |
| Network Model Categories                                                                  | 19 |
| The Process of Creating a Model                                                         | 19 |
| Issues Related to the Construction of Appropriate Measures Used to Calculate Entropy    | 20 |
| Entropy of a Set of Observations of a Study Variable                                     | 20 |
| Precision, Accuracy, and Validity of Study Data Considerations When Applying Conant’s Transmission Calculations to the Decomposition of Complex Systems | 20 |
| Accuracy of Entropy Calculations                                                         | 21 |
| Entropy Is Non-Directional                                                                | 21 |
| Summary                                                                                  | 21 |

## 4 METHODS ..................................................................................................................24

| Determination of the TCBH                                                               | 24 |
| Determination of a 24-Hour Model Instance                                               | 25 |
| Determination of the Four 6-Hour Model Instances                                        | 25 |
| Calculation of Conant’s Transmission Parameter                                           | 25 |
| Computing Model Parameters                                                               | 32 |
| Verification and Validation of Models                                                   | 35 |
| Comparison of Models                                                                    | 36 |
| Summary                                                                                  | 36 |

## 5 RESULTS ..................................................................................................................38

| Message Trunk TCBH Model Instance Versus 24-Hour Model Instance                         | 38 |
| ATM - HER Circuit TCBH Model Instance Versus 24-Hour Model Instance                    | 44 |
| Peering Exchange Router – Bilateral Peering Router Circuit TCBH Model Instance Versus 24-Hour Model Instance | 50 |
| ATM - HER Circuit 6-Hour Busy Period Model Versus TCBH and 24-Hour Models              | 56 |
# TABLE OF CONTENTS (Continued)

## CHAPTER

<table>
<thead>
<tr>
<th>Summary</th>
<th>58</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 DISCUSSION</td>
<td>59</td>
</tr>
<tr>
<td>Complex Communications Network Modeling and Decomposition by Entropic Techniques</td>
<td>59</td>
</tr>
<tr>
<td>An Exploration of Other Applications of Decomposition by Entropic Techniques</td>
<td>60</td>
</tr>
<tr>
<td>Service-based Entropic Model for Sensors Orchestration</td>
<td>60</td>
</tr>
<tr>
<td>Mapping Networks of Terrorist Cells</td>
<td>61</td>
</tr>
<tr>
<td>Summary</td>
<td>61</td>
</tr>
<tr>
<td>7 CONCLUSION</td>
<td>63</td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>65</td>
</tr>
</tbody>
</table>

## APPENDIX

<p>| A JAVA PROGRAMS FOR CALCULATIONS | 68 |
| B SAMPLE SPREADSHEET FOR 6-HOUR MODELS | 86 |
| C SAMPLE VOICE DATA | 91 |
| D SAMPLE ATM-HER DATA | 98 |
| E SAMPLE PXR-BILATERAL PEER DATA | 105 |</p>
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Examples of Entropy Values for Various Degrees of Busy Hour Coincidence</td>
</tr>
<tr>
<td>2</td>
<td>Message Trunk Mutual Information (Transmission) Results</td>
</tr>
<tr>
<td>3</td>
<td>Message Trunk Combination Model Prediction Results</td>
</tr>
<tr>
<td>4</td>
<td>ATM-HER Circuit Mutual Information (Transmission) Results</td>
</tr>
<tr>
<td>5</td>
<td>ATM-HER Circuit Combination Model Prediction Results</td>
</tr>
<tr>
<td>6</td>
<td>PXR Circuit Mutual Information (Transmission) Results</td>
</tr>
<tr>
<td>7</td>
<td>PXR Circuit Combination Model Prediction Results</td>
</tr>
<tr>
<td>8</td>
<td>ATM-HER Circuit 6-Hour Period Mutual Information (Transmission) Results</td>
</tr>
<tr>
<td>9</td>
<td>ATM-HER Circuit Combination 6-Hour Period Model Prediction Results</td>
</tr>
<tr>
<td>10</td>
<td>ATM-HER Combination Circuit Mutual Information (Transmission) Results</td>
</tr>
<tr>
<td>11</td>
<td>ATM-HER Circuit Combination Circuit Model Accuracy Comparison</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Some typical trunking arrangements</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Simplified Internet network diagram</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>ATM circuit 4 usage by hour of the day</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>Trunk 1 usage over time</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>Trunk 2 usage over time</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>Trunk 1 usage by hour of the day</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>Trunk 2 usage by hour of the day</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>Usage combination of two trunk groups</td>
<td>42</td>
</tr>
<tr>
<td>9</td>
<td>Trunk 1 plus 2 usage over time</td>
<td>42</td>
</tr>
<tr>
<td>10</td>
<td>Trunk 1 plus 2 usage by hour of the day</td>
<td>43</td>
</tr>
<tr>
<td>11</td>
<td>ATM circuit 1 usage over time</td>
<td>45</td>
</tr>
<tr>
<td>12</td>
<td>ATM circuit 2 usage over time</td>
<td>45</td>
</tr>
<tr>
<td>13</td>
<td>ATM circuit 1 usage by hour of the day</td>
<td>46</td>
</tr>
<tr>
<td>14</td>
<td>ATM circuit 2 usage by hour of the day</td>
<td>47</td>
</tr>
<tr>
<td>15</td>
<td>Usage combination of two circuits</td>
<td>48</td>
</tr>
<tr>
<td>16</td>
<td>ATM circuit 1 plus 2 usage over time</td>
<td>48</td>
</tr>
<tr>
<td>17</td>
<td>ATM circuit 1 plus 2 usage by hour of the day</td>
<td>49</td>
</tr>
<tr>
<td>18</td>
<td>PXR to Bilateral Peer circuit 1 usage over time</td>
<td>51</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>19</td>
<td>PXR to Bilateral Peer circuit 2 usage over time</td>
<td>51</td>
</tr>
<tr>
<td>20</td>
<td>PXR to Bilateral Peer circuit 1 usage by hour of the day</td>
<td>52</td>
</tr>
<tr>
<td>21</td>
<td>PXR to Bilateral Peer circuit 2 usage by hour of the day</td>
<td>53</td>
</tr>
<tr>
<td>22</td>
<td>Usage combination of two circuits</td>
<td>54</td>
</tr>
<tr>
<td>23</td>
<td>PXR to Bilateral Peer circuit 1 plus 2 usage over time</td>
<td>54</td>
</tr>
<tr>
<td>24</td>
<td>PXR to Bilateral Peer circuit 1 plus 2 usage by hour of the day</td>
<td>55</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Today’s voice and data connectivity typically utilizes a complex system of networks of networks. The development of models for characterizing usage patterns in complex networks would be useful in projecting capacity requirements in growing systems. The simplified models applied to a complex network as a timing and sizing algorithm would significantly reduce the amount of computations and data storage necessary to produce forecasts of capacity requirements. This composition introduces a network usage pattern model based on decomposition-by-information-transfer and verifies its application in real-life voice-data networks.

Problem Statement

Telecommunications and data networks are large-scale, complex engineering systems that require significant capital investment to establish and grow. The successful management of these networks can be simplified by decomposing these complex engineering systems into smaller sub-systems that can be economically modeled. One modeling issue is the task of predicting the impact of combining multiple traffic flows through a common device. This dissertation shows how information theory can be applied to decompose real-life complex communications systems to identify their relevant sub-systems so that timely, reliable, and economical architectures can be engineered through the use of models.
Importance

Information theory using communications as a model for describing natural phenomena can provide additional insight in understanding the behavior of complex engineering systems. Theoretical models of nature are generally limited in some way. For example, classical physics describes the behavior of ordinary-sized objects that move at ordinary velocities. For sub-atomic sized objects it is necessary to apply quantum physics. For high-velocity objects that travel at speeds approaching light, relativity physics are required. Relativistic quantum physics are necessary to model very fast and very small objects. Quantum theory can describe an object as a particle or as a wave but not both at the same time. Information can be transferred as energy from one point to another just as a particle or a wave. Information theory can be used to model objects using communications as a basis for describing phenomena.

Hypothesis

The connections among specific elements of a voice or data communications network have typical patterns of usage that contain information. The information content in the usage patterns can be used as a signature to identify the sub-systems of the overall network. Knowledge of these sub-systems can be used to plan appropriate architectures and economic capacity additions to ensure adequate service levels. Information theoretical analysis such as the calculation of transmission parameters among connections is one approach that can be used to decompose complex communication networks.
There is a wealth of published material that addresses information theory. This research is based on Shannon’s “A Mathematical Theory of Communication” [1], Ashby’s *An Introduction to Cybernetics* [2], Simon’s “The Architecture of Complexity” [3], and Conant’s “Detecting Subsystems of a Complex System” [4], as well as numerous other texts and articles that deal with probability, systems, telephony, the Internet, and complexity. Shannon discusses entropy as a measure of the information produced when a message is chosen from a finite set of possible messages [1]. Ashby provides numerous descriptions and examples of measures of variety and its transmission through a system. He shows us how entropy is a refinement of variety [2]. Simon describes how stable complex systems are usually composed of subsystems. The decomposition of complex systems is useful for the simplification of analysis and improvement of understanding [3]. Conant gives us a mathematical technique for decomposing complex systems based on the entropy of measurements of system parameters [4]. Ziemer and Tranter provide a basic explanation of information theory and concepts that are explained in greater detail by Sklar [5],[6].

Papoulis and Pillai and also Kreysig explain the mathematical theory used in the analysis of random variables. This provides us with the techniques necessary for determining probabilities used in the calculation of entropy [7],[8]. Linear algebra and associated matrix computations are a part of alternative decomposition techniques and are explained by Carlen and Carvalho and by Golub and Van Loan, respectively [9],[10].

Information entropy has close ties to thermodynamic entropy and the analysis of complex systems. A background in understanding basic physics, modern physics, and
thermodynamics is available in texts by Resnick and Halliday [11], Weidner and Sells [12], and Sonntag and Van Wylen [13]. Landauer describes the physical characteristics of information and relates information theory to underlying concepts from physics [14].

One of the purposes of modeling complex systems is to predict future behavior. A review of 25 years of research in time series forecasting is provided by De Gooijer and Hyndman [15]. Mihram [16] describes the categorization of models into 24 types based on the dimensions of abstraction and generality. He also describes approaches to the verification and validation of models [17].

Bell Communications Research published an industry standard text on trunk traffic engineering that provides the definitions and calculation methods for the analysis of message trunk examples [18]. Examples of these calculations, as well as more recent applications, can be found in Parkinson [19].

There are a number of references that address metrics associated with the Internet. Molyneux and Williams provide an overview of the many ways in which the Internet can be measured, from traffic volumes to numbers of web pages [20]. Prasad, Murray, Dovrolis, and Claffy define the metrics, techniques, and tools for estimating data bandwidth [21]. One of the fundamental needs of research is to find sources for relevant data, and an excellent catalog is provided by Shannon, Moore, Keys, Fomenkov, Huffaker, and Claffy [22]. Infrastructure for the collection of Internet measurements is described by Paxson, Mahdavi, Adams, and Mathis [23]. Issues associated with the forecasting or prediction of short-term behavior of Internet data flows is provided in articles by Hinich and Molyneux [24], and Zhang, Breslau, Paxson, and Shenker [25].
Paxson extends this concept with an article that explains the application of short-term projections to the identification of network intrusion events [26].

Barabási and Bonabeau [27] investigate network structures and describe scale-free characteristics, such as networks consisting of many nodes with few connections and a few nodes with many connections. Random networks would have nodes with the number of connections per node following a Poisson distribution. The Internet and the telephone network are not random networks and have scale-free characteristics.

Complex systems exist in many forms, not just in communications networks. The analysis of many types of networks and an approach for the identification of missing components is described by Clauset, Moore, and Newman [28]. An application of decomposition by transmission used to manage sensor networks is explained by Sadasivam, Goli, Kathiru, Krishnan, Thompson, Tuncer, and Tanik [29]. An anti-terrorist application of this research is documented by Krebs [30].

Summary

The successful management of complex communications networks can be simplified by decomposing complex engineering systems into smaller sub-systems that can be economically modeled. Information theory can be used to model objects using communications as a basis for describing phenomena. Information theoretical analysis, such as the calculation of transmission parameters among connections, is one approach that can be used to decompose complex communication networks. This study is based on research and publications in the areas of information theory, mathematics, physics, forecasting, modeling, traffic engineering, Internet metrics, and network structures.
CHAPTER 2

IMPROVING THE ECONOMIC DESIGN OF DATA AND VOICE NETWORK ARCHITECTURES USING INFORMATION THEORY

Economic Design of Voice Network Architectures

Current Approach Using Time-Consistent Busy Hour (TCBH)

The network design engineer is responsible for specifying an economic configuration and the number of transmission paths needed to meet the service objectives of the voice telephone system. The flow of traffic can be modeled using the principles of probability and statistics and the assumption that not all customers will attempt to use the network at the same time. Economic design involves the specification of timing, sizing, configuration, and technology selection of network components so that the expected flow of traffic will be served by a system capacity that is adequate to meet service objectives [18].

The establishment of appropriate service objectives is dependent on the technical characteristics of the serving component, customer expectations, regulatory requirements, and cost factors. For transmission paths, or message trunks, the service objectives are usually stated in terms of the probability of blocking a call or the average waiting interval to obtain a circuit. The methods for determining the service measurement must also be specified. A variety of message trunk configurations exist in the telephone network (Fig. 1).
Current design methods use the following criteria:

1. Criteria apply to the busy season, busy hour.

2. The weekday base period consists of 20 average business days in the busy season.

3. Weekends and holidays are excluded due to their different load profiles.

4. Hourly data, with timing that begins on the hour, is averaged for the record period.

5. Alternate paths for the same call are designed together as a cluster using the TCBH for the entire cluster.

6. Where only a single path is available for a call, the path is designed using the TCBH for that path.
One hour was chosen as the base unit to balance the cost of data collection and processing with required statistical accuracy.

The definition of a TCBH produces a single number model that represents an entire year’s traffic flow. This concise model, along with its definitions and constraints, can be used efficiently in calculations to size required network capacity. If the real world traffic pattern is consistent in time and across the entire path between the calling and called party, then a single number model can be a good representation. However, if the busy hour varies significantly from day to day or there are multiple segments in the path with non-coincident busy hours, then the single-hour model may be less accurate.

Suppose the busiest hour of the day fluctuates between two different hours among the 20 days in the record period. If the average of this “bouncing busy hour” was used instead of the TCBH, then the single-hour model would have a larger value and therefore indicate a need for more capacity to meet the same service objective.

In the configuration where a path consists of two or more segments in series, the design objectives must consider the additive effect of blocking in each segment. The single-hour model assumes all segments have coincident busy hours. If their actual busy hours are non-coincident, then the model would overstate the capacity required for a given service objective.

**Potential Improvement Using Time-Consistent Multi-hour (TCMH)**

An alternative model using the same criteria and assumptions as the single-hour model except for using the averages of each of the 24 daily hours may be more useful. The process for obtaining the 24 hourly numbers is the same as the single-hour method
except that all 24 hour average values are used in the model. Using this model, one can distinguish between loads that have a strong peak for one hour per day and those that have broader or multiple busy periods. This additional information is useful when the traffic loads from multiple paths are combined in some common element or multiple segments of the same path have different busy period patterns. The trade off for this additional information is additional computational and storage load.

Economic Design of Data Network Architectures

Current Approach of Over-Sizing Initially and Rebuild at Exhaust

The initial sizing of a data network architecture (Fig. 2) is based on a number of factors, including available components, manufacturer’s recommendations, projections of usage, field trials, and past experience. This usually results in more than adequate capacity at the turn up of the system with declining margins as network usage grows over time. The system performance is monitored by the network operator, and network capacity is added or the system architecture reconfigured to provide relief when performance degrades or is projected to degrade. This approach works well in an environment where there is rapid technological change or where capacity additions can be completed at very short intervals. In a rapidly evolving technological environment, new functionality may economically trigger a network replacement before capacity exhaust is approached. As a technology stabilizes or where networks grow so large that complete replacement becomes less practical, alternative approaches may be more effective.
Fig. 2. Simplified Internet network diagram.

*Potential Improvements by Projecting Infrastructure Exhaust Timing Using Sizing Models*

A simple data network could be described as two nodes connected by a link. This could be a client and server connected by a data path. A large-scale network would have many clients and many servers interconnected via multiple paths via numerous switches, routers, gateways, and other devices. Typical capacity constraints include the bandwidth of the links and interfaces, and the processing, memory, and storage capacity at the end and intermediate nodes. One approach to the design of capacity relief strategies follows.
First, the architecture is described as a model. The capacity constraints are identified and time series measurements collected for each component. This data is then trended and overlaid with market forecasts to identify the timing of exhaust points and the sizing of relief projects. The compilation of capital relief projects, and revenue and expense projections, all play a part in the development of a business plan and of annual capital and expense budgets.

For large-scale networks, the tasks of collection and trending of data are major undertakings. The data collected may originate from performance monitoring by the network operator. Performance monitoring requires a large volume of very granular (typically sub-minute) data that is kept for a relatively short period of time and then discarded. Performance monitoring is used to identify failed or overloaded components or software that can be replaced, bypassed or reconfigured to maintain objective service levels. Abnormal conditions can be identified by comparing current measurements to prior hours, days or weeks data. The development of a long-term relief plan requires a longer time horizon but less granular data. The performance monitoring data can be aggregated and averaged over longer intervals, such as hourly periods. Even at this level of abstraction, the data storage and processing requirements for large-scale networks is very costly. This makes the use of simplified data models essential to the capacity and architecture design process.

One could use a similar approach to voice telephony and model traffic usage based on a TCBH. Telephony traffic is dominated by either business use with daytime busy hours or residential use with evening busy hours. On the other hand, data traffic often has more complex patterns. This makes the use of more complex models a valuable
consideration. Where there is a choice among models there is a need for objective
criteria to guide that decision. A criterion that is numerical in nature can be more readily
utilized in a mechanized system than an abstract or visual interpretation of graphical
plots.

*Other Approaches*

A bouncing busy hour model is similar to the TCBH model except that for each
workday the hour with the highest usage is selected. The average of these “bouncing”
busy hours is then used in the sizing and time algorithms. The advantage is a
simplification in the calculation of the busy hour average. This average will always be
greater than or equal to the TCBH average. If the sizing and timing criteria are not
adjusted then over-sizing or early timing of capacity may result.

Peak value models can be used for timing and sizing network elements where any
demand greater than the capacity is considered a failure. These models are particularly
sensitive to any extreme data points, and the process for excluding erroneous data must
be rigorous.

Models may be constructed that use a count of threshold crossings as the usage
measure. These models can be used to transform data sets that include wide ranges of
usage values where average usage would otherwise be dominated by relatively few large
data values. A telephony example would be the use of “peg counts” where each call is
counted without regard to its length.
Summary

The network design engineer is responsible for specifying an economic configuration and a number of transmission paths to meet the service objectives of the voice telephone system. The definition of a TCBH produces a single number model that represents an entire year’s traffic flow. This concise model, along with its definitions and constraints, can be used efficiently in the calculations needed to size the required network capacity. An alternative model using the same criteria and assumptions as the single-hour model except for using the averages of each of the 24 daily hours may be more useful.

The initial sizing of data network architecture is based on many factors, including available components, manufacturer’s recommendations, projections of usage, field trials, and past experience. This usually results in more than adequate capacity at the turn up of the system with declining margins as network usage grows over time. One could use a similar approach to voice telephony and model traffic usage based on a TCBH. Telephony traffic is dominated by either business use with daytime busy hours or residential use with evening busy hours. On the other hand, data traffic often has more complex patterns. This makes the use of more complex models a valuable consideration.

Knowledge of the underlying processes and the purpose of the study are essential in the selection of appropriate models.
CHAPTER 3
MODELING TRAFFIC USAGE

Comparing the Activity of the Elements of a Complex System to Simplified Models of the System Using Information Contained in the Measures

*Measurement Data May Contain Structure*

If measurement data is not random, then there will be some structure that can be leveraged to create a simplified model. In voice telephony there is typically a busy season lasting several months during which the average usage is greater than other months. Weekdays are usually busier than weekends. During the day, usage is highest during business hours in a commercial environment and highest in the evenings in a residential environment. Data circuits may follow these patterns or they may have other patterns, such as large file transfers during overnight hours or monthly software updates that are sent to a large number of clients.

*Structure Can Be Modeled*

In the case of voice telephony, the traditional model is the TCBH. For other systems, such as data networks, a 24-hour model may be more appropriate. Current practice for data networks is to model a circuit by its peak usage over a recent time interval.
Models Can Be Compared to Observed Data Using Entropic Measures

Conant used an entropic measure he described as transmission to identify subsystems within a complex system [4]. In a similar fashion, a data set can be combined with its model, and the transmission parameter between the model and the data set then calculated, to measure the strength of the model in describing the associated data set.

Entropic Measures

Conant’s method involves the calculation of the transfer of information among components of a subsystem. The transmission measure is based on the information content of each part of the subsystem and the mutual information between or among parts. One of the first requirements for a measure of information transfer is for the metric to contain some variety. As defined by Ashby [2], variety is the number of distinguishable elements. If a time series of measures had a constant value, then the information content would be zero.

Papoulis [7] defines mutual information, $I(U, B)$, as follows: Assign a measure of uncertainty to a partition of mutually exclusive events of an experiment. “For any partitions $U$ and $B$, the function $I(U, B) = H(U) + H(B) - H(U \cdot B)$ is called the mutual information of the partitions $U$ and $B$.” The entropies of the partitions, $U$, $B$, and the product of their partitions, $U \cdot B$, are denoted as $H(U)$, $H(B)$, and $H(U \cdot B)$. He defines the product of two partitions as follows: “The product of two partitions $U = [A_i]$ and $B = [B_j]$ is a partition whose elements are all intersections $A_i \cap B_j$ of the elements of $U$ and $B$. This partition will be denoted by $U \cdot B$. Clearly, $U \cdot B$ is the largest common refinement of $U$ and $B$. ”
The definition of entropy, $H(U)$, when used as a measure of information content can be described in four ways [7]:

- **Subjective**: $H(U)$ is a measure of uncertainty about the occurrence of the events $A_i$ of the partition $U$ in a single performance of experiment $S$.

- **Principle of maximum entropy, (ME)**: The probabilities $p_i = P(A_i)$ must be such as to maximize $H(U)$ subject to the given constraints. Since the number of typical sequences, $n_t = e^{nH(U)}$, where $n$ is the number of repeated trials, the ME principle is equivalent to the principle of maximizing the number of typical sequences. If nothing is known about the probabilities $p_i$, then the ME principle leads to the estimates $p_i = 1/N$, $H(U) = \log N$, and $n_t = N^n$, where $N$ is the number of elements in the partition.

- **Axiomatic**: $H(U)$ is a number assigned to each partition of $S$. This number equals the sum, $-\sum p_i \log p_i$, where $p_i = p(A_i)$.

- **Empirical**: This involves the repeated performance of the experiment $S^n$ of repeated trials. A specific typical sequence $t_j$ is an event with probability $e^{-nH(U)}$. Using the relative frequency interpretation of probability, we conclude that if the experiment $S^n$ is repeated $m$ times and the event $t_j$ occurs $m_j$ times, then for sufficiently large $m$,

$$P(t_j) = e^{-nH(U)} \approx \frac{m_j}{m}, \text{ hence } H(U) \approx -\frac{1}{n} \ln \frac{m_j}{m}.$$  

This relates the theoretical quantity $H(U)$ to the experimental numbers $m_j$ and $m$. 
Except for the subjective interpretation, all involve the calculation or estimation of probability.

**Creation of Simplified Models of Internet Circuit Usage Patterns Using Entropic Measures**

*Examples of Internet Circuit Usage Measurement Data Contain Structure*

Scatter plots can be an effective visual tool to identify daily usage patterns. When hourly data is plotted versus time of day, dominant usage periods may stand out. A scatter plot of circuit usage versus hour of the day (Fig. 3) illustrates that usage is low during late night and early morning hours and higher during business hours and evening hours. This plot also shows two distinct usage peaks, one at 2 p.m. and one at 7 p.m.

*Creating and Comparing Various Models of Internet Circuit Usage Using Entropic Measures*

The data can be structured into a 2-state model by defining a threshold and grouping the data points into busy or non-busy states. This is similar to the TCBH model but could also include models with multiple hours. Another approach is to define a multi-state model with multiple thresholds, such as a 24-hour model with a state defined for each hour of the day. A model can be developed that groups multiple hours to obtain more information than the single busy hour model but requires less data than a 24-hour model. These multi-hour models can be arbitrary groups of hours or hours selected based on their individual average usage over time.
Development and Validation of Models

Categories of Models

Mihram [16] describes modeling as “the capability to describe large-scale complicated interactive systems by symbolic representations so that inferences regarding the effects of alternative system configurations can be easily and rapidly structured.” He proceeds to categorize models as material or symbolic along one dimension and static or dynamic along a second dimension. These categories are then further refined by dividing the material category into the sub-categories of replication, quasi-replica, and analogue. The symbolic category is divided into the sub-categories of descriptive, similar, and formalization. As one moves along this dimension from material-replication to symbolic-formalization the model characteristics increase in abstraction and inferential facility and
decrease in reality. Along the other dimension, both the static and dynamic categories are divided into deterministic and stochastic sub-categories. As one moves from static-deterministic to dynamic-stochastic, generality increases. This produces a total of 24 sub-categories of model types.

Network Model Categories

There are three models described in this dissertation: a TCBH model, a 6-hour busy hour model, and a 24-hour model. All three of these are symbolic of the systems they represent and have a formalized definition. The models are dynamic in that their results are dependant upon time of day and day of week.

The Process of Creating a Model

Mihram [16],[17] also describes a step-by-step process for constructing models based on the builder’s a priori knowledge of the systems to be studied. Before the model is constructed, its purpose should be clearly defined and a process for measuring its effectiveness determined.

The first step is systems analysis. Here the behavior of the system and the known relationships of its components and their behaviors and interactions are identified. This is an appropriate time to determine the category of model since the success of the modeling effort may depend on a good match of model category to the system under study and the purpose of the model.

Step two is system synthesis. In this stage the preliminary model is built or programmed based on the preceding analysis.
The third step is verification that the model performs according to its specifications. For a material model this may involve checking that dimensions are scaled correctly. For a computer program model this is the debugging phase.

In step four, the model is validated against a real system to determine if the model produces a credible representation of the system that is under study.

The inference step is when the model is put to its intended use as a representative of the real object or system.

Issues Related to the Construction of Appropriate Measures Used to Calculate Entropy of a Set of Observations of a Study Variable

*Precision, Accuracy, and Validity of Study Data Considerations When Applying Conant’s Transmission Calculations to the Decomposition of Complex Systems*

The transformation of raw data into useful information is an important step in any study involving measurement data. Parameters of interest must be identified and raw data, which may be in a different form, transformed to be relevant to the analysis. Performance monitoring data may be collected in one-second intervals, but the study may require hourly averages or peaks or counts of threshold crossings. The analyst must decide what constitutes a relevant state, such as busy versus non-busy. There may be multiple relevant states to be used in the study that must be defined.

In order for the measure to be useful it must be valid. A measurement data point can be true to its actual value, in other words, accurate. The measurement can be repeatable with sufficient precision. But accuracy and precision are not sufficient alone. It is important for the analyst to use domain expertise to ensure that the results truly represent the concept to be measured.
Accuracy of Entropy Calculations

A large number of observations is often required for an accurate calculation of the entropy of a process, since the entropy is estimated by counting the number of occurrences of each possible process state. The number of observations required for a given degree of accuracy is dependent on the variety of the measure. For example, the sample size must far exceed the number of possible values of the measure in order to calculate the probability of each occurrence of each value. This could present problems for models with more than two possible states. However, in the comparison of a model to a set of observations, one is not concerned with estimating the probabilities of the underlying process but rather the probabilities of each value in the particular set of observations.

Entropy Is Non-Directional

Models based on entropy calculations must be carefully constructed in order to avoid confusing a pattern with its inverse. For example, a usage pattern with a single high hour and 23 low hours will have the same entropy as a usage pattern with a single low hour and 23 high hours but the effect on a network system is distinctly different.

Summary

If measurement data is not random, then there will be some structure that can be leveraged to create a simplified model. In the case of voice telephony, the traditional model is the TCBH. For other systems, such as data networks, a 24-hour model may be more appropriate. A data set can be combined with its model, and Conant’s transmission
parameter then calculated between the model and the data set, to measure the strength of the model in describing the associated data set.

Scatter plots can be an effective visual tool to identify daily usage patterns. When hourly data is plotted versus time of day, dominant usage periods may stand out.

Various models of Internet circuit usage can be created and compared using entropic measures. The data can be structured into a two-state model by defining a threshold and grouping the data points into busy or non-busy states. Another approach is to define a multi-state model with multiple thresholds, such as a 24-hour model with a state defined for each hour of the day.

Models can be categorized across two dimensions as material or symbolic along one dimension and static or dynamic along a second dimension. Network models in this paper are symbolic and dynamic. There is a step-by-step process defined for the construction of models.

There are issues related to the construction of the appropriate measures to use when calculating the entropy of a set of observations of a study variable. Precision, accuracy, and validity of study data should be considered when Conant’s transmission calculations are applied to the decomposition of complex systems. It is important for the analyst to use domain expertise to ensure that the results truly represent the concept to be measured. A large number of observations are often required for an accurate calculation of the entropy of a process. However, in the comparison of a model to a set of observations, one is not concerned with estimating the probabilities of the underlying process but rather the probabilities of each value in the particular set of observations.
Entropy is non-directional, so models based on entropy calculations must be carefully constructed to avoid confusing a pattern with its inverse.
CHAPTER 4

METHODS

This chapter lists the details of the design for three types of model applications: TCBH, 6-hour busy period, and a 24-hour model application. Conant’s decomposition by transmission is described with examples. Step-by-step details for calculating the Conant transmission parameters for the study model applications are specified. An approach to validating the model is defined, and the method for comparative evaluation of models described.

Determination of the TCBH

The process for determining the busy hour involves the following steps:

1. Collect hourly traffic data for all available days during the busy season.
2. Exclude weekend and holiday hours from the data.
3. For each of the 24 hours in a day, sum the busy season usage and divide by the number of hours included in the sum to obtain the average.
4. The busy hour is the one with the greatest average value.

A Java program was written that takes a set of hourly measurements as input and computes the TCBH and its average value for the data set using these steps. Pseudo-code for the program is listed in Appendix A.
Determination of a 24-Hour Model Instance

The process for determining a 24-hour model is similar to the TCBH model and involves the following steps:

1. Collect hourly traffic data for all available days during the busy season.
2. Exclude weekend and holiday hours from the data.
3. For each of the 24 hours in a day, sum the busy season usage and divide by the number of hours included in the sum to obtain the average.
4. The model is defined by the 24 hourly averages.

The Java program in Appendix A also computes the 24-hour model and its hourly averages using these steps.

Determination of the Four 6-Hour Model Instances

Model “B” is defined as follows: the six hours beginning at midnight are busy, and the other eighteen hours are not busy. Model “C” is defined as follows: the six hours beginning at 7 a.m. are busy, and the other eighteen hours are not busy. Model “D” is defined as follows: the six hours beginning at noon are busy, and the other eighteen hours are not busy. Model “E” is defined as follows: the six hours beginning at 6 p.m. are busy, and the other eighteen hours are not busy. This model was developed using a spreadsheet, and a sample of a worksheet is shown in Appendix B.

Calculation of Conant’s Transmission Parameter

Conant’s method involves the calculation of the transfer of information among the components of a subsystem. The transmission measure is based on the information
content of each part of the subsystem and the mutual information between or among parts. Determining transmission parameters involves the calculation of probabilities, 

\[ P(A) \]

Use of the relative frequency definition \( P(A) = \lim_{n \to \infty} \frac{n_A}{n} \) requires counting occurrences: \( n_A, n, \text{ and } A \). What does \( A \) mean? In a set of observations of an experimental variable, \( A \) represents the possible states of the variable. If the variable is not metric then \( A \) represents the possible categories. For example, if the observations were of a traffic signal, then possibilities could include such values as red, yellow, and green. If the variable is a continuous metric or if the number of possible discrete values is large, then ranges of values can be used to limit the number of states. The selection of states involves knowledge of the process under study. In an appropriate selection, the difference from one state to another should have meaning in the context of the study.

What do \( n \) and \( n_A \) mean? In general, \( n \) should equal the number of valid observations for which states \( A \) are possible outcomes. The number of observations with state \( A \) equals \( n_A \).

Counting occurrences using a program can be accomplished as follows:

1. Establishing a counter for each possible value of \( A \),

2. Comparing each observation to the possible values of \( A \) and increment the appropriate counter.

An alternate approach may be more efficient where there are a large number of possible values of \( A \) compared the actual number of values of \( A \) that occur.

1. First, sort the observations by the values of \( A \).

2. Establish a counter for the first \( A \) value in the list.
3. Read the next observation in the list and increment the counter if the value of $A$ matches the previous observation.

4. If the next observation does not match the previous, then create a new counter and increment as in the previous step.

5. Once all the observations in the list have been read, the set of counters will represent the sub-set of values of $A$ that actually occurred.

Entropy calculations involving the intersections of sets of observations are the most challenging. If the sets of variables involve sequential time series observations, then the intersection of sets requires timing alignment among sets. Suppose the first set has an observation at midnight of January 1, 2008, and subsequent observations each hour for 60 days. Further suppose that the second set has its first observation on January 2, 2008, at 1 a.m., and with subsequent observations each hour for 60 days. In order to align these two sets of data the first 25 observations in set one and the last 25 observations from set two would be discarded. Missing or invalid values could be replaced by interpolated values and documented. The resulting union of sets would begin with the combination of the value of set one on January 2, 2008, at 1 a.m., with the value of set two on January 2, 2008, at 1 a.m., and continue with each subsequent value determined by combining subsequent observations from set one with subsequent observations from set two. If a time offset of one hour is desired in the union, then the first combination would consist of the January 2, 2008, at midnight observation from set one combined with the January 2, 2008, at 1 a.m., observation from set two.

If set one has a number of states equal to $X$ and set two has a number of states equal to $Y$ then the combination could have as many states as $X$ times $Y$. The possible number of
combined states can become large if multiple sets with multiple states are combined. Counting the occurrences of the intersection of sets can be accomplished in a manner similar to counting occurrences of a single set, with the parameter $A$ equal to the number of combined states.

The steps to calculate the intensity measure normalized transmission, $t_{ij}$, are summarized as follows [4]:

1. Consider a complex system that has $K$ primary variables, each of which is observed once every standard time increment $N$, giving a total of $K \cdot N$ observations. With each of the primary variables is associated a derived variable $X_j, 1 \leq j \leq K$ that can be grouped into sets.

2. If the $K$ primary variables are not integer metrics then derived integer variables, $X_j, 1 \leq j \leq K$, are chosen to represent the subsystem categories.

3. The variables are grouped into sets of time series observations.

4. The number of occurrences of each possible state value in each set is counted and used to calculate the entropy of that set by

$$H(X_j) = \log_2 N - \frac{1}{N} \sum_{i=1}^{M_j} n_i \log_2 n_i$$  \hspace{1cm} (1)

where $j$ is the state, $H(X_j)$ represents the entropy of the variable $X_j$, $N$ is the total number of observations, $M_j$ is the upper limit of the range of values of $X_j$, $i = \text{index value}$, and $n_i = \text{number of observations of each index value}$.

5. Vectors representing pair-wise combinations of each set of a variable’s time series observations with the second variable offset by one time step are generated.
6. The number of occurrences of each possible combination of values in each vector is counted for each vector and used to calculate the joint entropy $H(X_i, X'_j)$, where $X'_j$ represents the set of observations offset by one time increment.

7. The transmission parameter is then calculated by

$$T(X_i : X'_j) = H(X_i) + H(X'_j) - H(X_i, X'_j).$$

(2)

8. The normalized transmission, $t_g$, can be obtained by dividing $T(X_i : X'_j)$ values by $H(X'_j)$

$$t_g = \frac{T(X_i : X'_j)}{H(X'_j)}.$$  

(3)

A telephony example of information transfer using the axiomatic definition of entropy follows. Consider a set of message trunk groups with known usage characteristics. The number of trunks required to meet a specified grade of service is generally based on the average of the usage during the busy hour in each of the twenty highest days in each month during the busy season.

If a busy season lasts for 3 months, then there are 60 busy hours included. Assuming 30 day months, there is a total of 3 times 30 times 24 or 2160 total hours. The trunk groups will be in their busy hour 2.78% of the time and not in a busy hour 97.22% of the time.

If there is uncertainty about the outcome of an event, then entropy can be used as a measure of the amount of information associated with knowledge of the outcome. For an event with two possible outcomes, entropy ranges from a minimum of zero to a
maximum of one. The maximum occurs when each possible outcome is equally likely. If either outcome is certain then the entropy is zero.

If each hour is defined as belonging to one of two possible states, busy or not busy, then the entropy can be calculated (5) using (4) from the probabilities of each state without knowing the usage values for each hour as

\[ H = -P(busy) \log_2 P(busy) - P(not \ space busy) \log_2 P(not \ space busy) \]  \hspace{1cm} (4)

\[ H = -(0.0278) \log_2(0.0278) - (0.9722) \log_2(0.9722) = 0.1832. \]  \hspace{1cm} (5)

This entropy value is based on the definition and is not based on the usage values of any individual trunk group. The entropy is small relative to the maximum possible, since over 97% of the hours are not busy hours.

When traffic from multiple trunk groups is combined on a network element, the usage is combined on a moment-by-moment basis such that if the busy hours are non-coincident then the resulting busy hour usage of the combination will be less than the sum of the two trunk groups at their individual busy hours. As an example, if trunk group A predominately serves a business community then its typical busy hour may be 10 a.m. If trunk group B serves a residential community then its typical busy hour may be 3 p.m. The capacity required for an element that serves both groups together would be smaller than the capacity required for each group individually.

Now look at this example from an entropic perspective. Both trunk groups have identical entropies based on the definition given above. The entropy of the combination can be defined in a different way considering that there are now four possible states instead of two. For any given hour, A and B can both be busy, A and B can both be not busy, A can be busy and B not busy, or A can be not busy while B is busy.
Look at extreme cases. Suppose A and B are always busy at the same times. If so, the combination will have the same entropy as A or B individually. The other extreme is more interesting; that is, when A and B are never busy at the same time. The probability that A is busy and B is busy, \( P(AB) \), is zero. The probability that A is busy and B is not busy, \( P(ABN) \), is 0.0278. The probability that A is not busy and B is busy, \( P(ANB) \), is 0.0278. The probability that A and B are both not busy, \( P(ANB) \), is 0.9444. The entropy (7) is calculated using (6) as

\[
H = -P(AB) \log_2 (P(AB)) - P(ABN) \log_2 (P(ABN)) - P(ANB) \log_2 (P(ANB)) - P(ANB) \log_2 (P(ANB))
\]

\[
H = 0 - (0.0278) \log_2 (0.0278) - (0.0278) \log_2 (0.0278) - (0.9444) \log_2 (0.9444) = 0.3653
\]

The entropy is greater in this case since the probability of both trunk groups being not busy is reduced.

Now consider an arbitrary point in between these extremes. If A and B coincide half of the time, the entropy may be calculated using (6) as

\[
H = -(0.0139) \log_2 (0.0139) - (0.9583) \log_2 (0.9583) = 0.3161
\]

The probability that A and B are both busy is 0.0139. The probability that A is busy and B is not busy is 0.0139. The probability that A is not busy and B is busy is 0.0139. The probability that A and B are both not busy is 0.9583. In this case there is less certainty (and therefore more entropy) than the coincident case that both trunk groups will be not busy. There is more certainty (and less entropy) than the non-coincident case.

Now consider Conant's approach to the identification of sub-systems of a complex system [4]. If usage from a number of trunk groups is combined in some element, then usage on that element could be modeled as a complex system. Two groups
with similar busy hours could be considered a sub-system, while groups with non-coincident busy hours could be considered separate systems. Examples of normalized transmission (10) through (14) are calculated using (9) as

\[
t = \frac{H(A)}{H(B)} + \frac{H(B)}{H(B)} - \frac{H(AB)}{H(B)}
\]

(9)

\[
t_{\text{coincident}} = \frac{H(0.1832)}{H(0.1832)} + \frac{H(0.1832)}{H(0.1832)} - \frac{H(0.1832)}{H(0.1832)} = 1 + 1 - 1 = 1
\]

(10)

\[
t_{75\%} = 2 - \frac{0.2653}{0.1832} = 0.5519
\]

(11)

\[
t_{50\%} = 2 - \frac{0.3161}{0.1832} = 0.2748
\]

(12)

\[
t_{25\%} = 2 - \frac{0.3511}{0.1832} = 0.0838
\]

(13)

\[
t_{\text{non-coincident}} = 2 - \frac{0.3653}{0.1832} = 0.0063.
\]

(14)

Table 1 lists corresponding entropies and normalized transmission values for several cases.

By using data with known relationships it has been shown that Conant’s approach can identify sub-systems within a complex system.

Computing Model Parameters

The following is the question to be answered. How well does a given model represent a real communications path? One approach is to determine the mutual information content between the model and the data to be represented. Intuitively, this information content would represent the amount of information that the model provides
Table 1

Examples of Entropy Values for Various Degrees of Busy Hour Coincidence

<table>
<thead>
<tr>
<th>Normalized Transmission $t$ Case</th>
<th>AB Entropy</th>
<th>A to B Normalized Transmission $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{\text{coincident}}$</td>
<td>0.1832</td>
<td>1.0000</td>
</tr>
<tr>
<td>A and B have coincident busy hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{75%}$</td>
<td>0.2653</td>
<td>0.5519</td>
</tr>
<tr>
<td>A and B have 75% coincident busy hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{50%}$</td>
<td>0.3161</td>
<td>0.2748</td>
</tr>
<tr>
<td>A and B have 50% coincident busy hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{25%}$</td>
<td>0.3511</td>
<td>0.0838</td>
</tr>
<tr>
<td>A and B have 25% coincident busy hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{\text{non-coincident}}$</td>
<td>0.3653</td>
<td>0.0063</td>
</tr>
<tr>
<td>A and B have non-coincident busy hours</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

about the data it represents. Mutual information is calculated as

$$I_{\text{mutual}} = H_{\text{model}} + H_{\text{data}} - H_{\text{model},\text{data}}.$$ (15)

The steps in computing the mutual information are as follows:

1. Determine the information content of the model by calculating its entropy. In the single-hour case, there are two possible states, busy or not busy. The busy hour is considered busy, and the other 23 hours are considered not busy. The single-hour entropy is computed as

$$H_{\text{single hour}} = -\frac{1}{24} \log_2 \left( \frac{1}{24} \right) - \frac{23}{24} \log_2 \left( \frac{23}{24} \right) = 0.19 + 0.06 = 0.25.$$ (16)

For the 24-hour model, one can define 24 different states and calculate the entropy as

$$H_{\text{24 hour}} = 24 \left( -\frac{1}{24} \log_2 \left( \frac{1}{24} \right) \right) = 4.58.$$ (17)
2. Determine the information content of the data by calculating its entropy. If there
are 20 weekdays in the study period then there are a total of 480 hours, 20 hours
(one each day) that would be considered busy and the remaining 460 not busy.
The entropy is calculated as

\[
H_{\text{data-single hour}} = -\frac{20}{480} \log_2 \left( \frac{20}{480} \right) - \frac{460}{480} \log_2 \left( \frac{460}{480} \right) = 0.19 + 0.06 = 0.25 .
\] (18)

For comparing the data to the 24-hour model, the data can be grouped into 24
states using percentiles producing the entropy as

\[
H_{\text{data-24 hour}} = 24 \left( -\frac{20}{480} \log_2 \left( \frac{20}{480} \right) \right) = 4.58 .
\] (19)

3. The entropy of the combinations of the models with the data is data-dependent.

For the single-hour model, each hour can be in one of four possible states, model
busy and data busy, model not busy and data not busy, model not busy and data
busy, and model busy and data not busy. The number of occurrences of each state
are summed and used in the entropy computation as

\[
H_{\text{model,data}} = -\frac{n_{\text{busy,busy}}}{480} \log_2 \left( \frac{n_{\text{busy,busy}}}{480} \right) -\frac{n_{\text{not busy,not busy}}}{480} \log_2 \left( \frac{n_{\text{not busy,not busy}}}{480} \right) -\frac{n_{\text{not busy,busy}}}{480} \log_2 \left( \frac{n_{\text{not busy,busy}}}{480} \right) -\frac{n_{\text{busy,not busy}}}{480} \log_2 \left( \frac{n_{\text{busy,not busy}}}{480} \right) .
\] (20)

If the data matched the model perfectly and every data busy hour coincided with
the model busy hour, then the last two terms would equal zero, and the entropy of
the combination would be equal to the entropy of the model. In the other
extreme, where the data busy hour never coincided with the model busy hour, the
first term is zero and the result equals about twice the entropy of the model. The
24-hour model has 576 possible states (24 squared). Fortunately, most of these
states do not occur in practical situations, and it is only necessary to count and calculate the entropy using the states that actually do occur.

4. The final step is to combine the entropies of the model and the data and subtract the entropy of the combination. The mutual information is at its maximum value (equal to the model entropy) when the data is exactly as predicted by the model. When comparing models with different entropies, it may be appropriate to normalize by dividing all terms of the equation by the model entropy to produce a normalized metric between zero and one.

The Java programs listed in Appendix A are used to compute the probabilities, entropies, and associated transmission parameters for the TCBH and for the 24-hour models using the approaches in the preceding sections. The spreadsheet program illustrated in Appendix B is used to compute the probabilities, entropies, and associated transmission parameters for the 6-hour busy period models using these same approaches.

Verification and Validation of Models

The computer models in this study were verified by using data with known characteristics and comparing the model output to the expected results.

The purpose of the models in this study is to predict the effects of combining traffic loads through common network elements. To determine the validity of the models, the traffic data were added together on an hour-by-hour basis and the resulting model usage of the combined traffic compared to the usage of the combinations of the models. As expected, models with higher potential information content had more accurate predictions than models with lower information content.
Comparison of Models

Once data has been collected transformed and modeled, the next step is to determine if the model is a good fit for the data. One approach to testing the validity of the model is to use combinations of data sets to see how well the combinations of their models predict the result. The steps are as follows:

1. Select sets of data with the same time period measurements.
2. Eliminate any non-overlapping data points.
3. Compute the models and their associated usage, busy and non-busy periods, and normalized transmission, $t$.
4. Combine the data sets on an hour-by-hour basis to form a new data set.
5. Compute the models for the new data set.
6. Combine the models from the original data sets.
7. Compare the usage of the combination of the original models to the usage of the model of the new data set.
8. The smaller the deviation between the combination of models and the model of the combined data, the better the model.

Summary

This chapter lists the details of the design for three types of model applications: TCBH, 6-hour busy period, and a 24-hour model application. Conant’s decomposition by transmission is described with examples. Step-by-step details for calculating the Conant transmission parameters for the study model applications are specified. An
approach to validating the models is defined. In addition, the method for comparative evaluation of models is described.
CHAPTER 5

RESULTS

Message Trunk TCBH Model Instance Versus 24-Hour Model Instance

This process has been tested by using a computer program to perform the calculations on sets of actual traffic data. The results of calculating the normalized transmission parameters and models for the TCBH and for 24 hours for usage in centum call seconds (CCS) on five different telephone message trunks show a greater transmission for the 24-hour models (Table 2). The normalized transmission, $t$, ranges from 0.002 to 0.007 for the TCBH model and from 0.381 to 0.439 for the 24-hour model. The TCBH for three of the five trunks is the hour beginning at 17:00 or 5 p.m. The other two trunks have busy hours beginning at 11 a.m.

Table 2

<table>
<thead>
<tr>
<th>Trunk</th>
<th>TCBH Model Normalized Transmission $t$</th>
<th>Busy Hour</th>
<th>24-Hour Model Normalized Transmission $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.002</td>
<td>17</td>
<td>0.389</td>
</tr>
<tr>
<td>2</td>
<td>0.007</td>
<td>17</td>
<td>0.439</td>
</tr>
<tr>
<td>3</td>
<td>0.006</td>
<td>17</td>
<td>0.435</td>
</tr>
<tr>
<td>4</td>
<td>0.007</td>
<td>11</td>
<td>0.437</td>
</tr>
<tr>
<td>5</td>
<td>0.003</td>
<td>11</td>
<td>0.381</td>
</tr>
</tbody>
</table>
Trunk usage as a function of time shows a distinct pattern with daily and weekly cycles (Figs. 4 and 5).

Fig. 4. Trunk 1 usage over time.

The same data when plotted as a scatter chart of usage versus hour of the day shows a pattern of low usage during late night and early morning hours with two patterns of higher usage from mid-morning to late evening (Figs. 6 and 7). The upper band of usage points is generally during workdays with the lower band during weekends.
Fig. 5. Trunk 2 usage over time.

Fig. 6. Trunk 1 usage by hour of the day.
The data from the five trunks was combined on an hour-by-hour basis for each of the ten possible combinations (Fig. 8). Four of the ten combinations included trunks with the same busy hour (Table 3). In those cases there was no perceptible difference in the accuracy of the two models in predicting the combined busy hour usage. For the six combinations of trunks with different busy hours, the 24-hour model was more accurate than the TCBH model by 0.8% to 3.5%.

The combination usage for Trunks 1 plus 2 is plotted over time (Fig. 9) and plotted as a scatter chart by hour of the day (Fig. 10). The usage combination charts for all the trunk combinations in the study are included in Appendix C.
Fig. 8. Usage combination of two trunk groups.

Fig. 9. Trunk 1 plus 2 usage over time.
Fig. 10. Trunk 1 plus 2 usage by hour of the day.

Table 3

Message Trunk Combination Model Prediction Results

<table>
<thead>
<tr>
<th>Trunks Combined</th>
<th>Combined Busy Hour Same Average Usage Estimate</th>
<th>TCBH Model Usage Estimate</th>
<th>24-Hour Model Usage Estimate</th>
<th>% Error</th>
<th>% Error</th>
<th>% Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 and 2</td>
<td>Yes</td>
<td>11859.2</td>
<td>11859.3</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>1 and 3</td>
<td>Yes</td>
<td>9981.7</td>
<td>9981.7</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>1 and 4</td>
<td>No</td>
<td>6345.4</td>
<td>6525.1</td>
<td>2.8%</td>
<td>0.9%</td>
<td>1.9%</td>
</tr>
<tr>
<td>1 and 5</td>
<td>No</td>
<td>4400.4</td>
<td>4488.2</td>
<td>2.0%</td>
<td>0%</td>
<td>2.0%</td>
</tr>
<tr>
<td>2 and 3</td>
<td>Yes</td>
<td>16380.6</td>
<td>16380.6</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>2 and 4</td>
<td>No</td>
<td>12574.6</td>
<td>12924.0</td>
<td>2.8%</td>
<td>1.6%</td>
<td>1.2%</td>
</tr>
<tr>
<td>2 and 5</td>
<td>No</td>
<td>10799.3</td>
<td>10887.1</td>
<td>0.8%</td>
<td>0%</td>
<td>0.8%</td>
</tr>
<tr>
<td>3 and 4</td>
<td>No</td>
<td>10676.1</td>
<td>11046.4</td>
<td>3.5%</td>
<td>0%</td>
<td>3.5%</td>
</tr>
<tr>
<td>3 and 5</td>
<td>No</td>
<td>8921.8</td>
<td>9009.5</td>
<td>1.0%</td>
<td>0%</td>
<td>1.0%</td>
</tr>
<tr>
<td>4 and 5</td>
<td>Yes</td>
<td>5553.6</td>
<td>5552.9</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>
ATM - HER Circuit TCBH Model Instance Versus 24-Hour Model Instance

Similar calculations on five Internet circuits that connect asynchronous transfer mode (ATM) access switches to high-speed edge routers (HER) show greater transmission values for the 24-hour models (Table 4). In this study the normalized transmission parameters, $t$, for the TCBH model ranged from 0.008 to 0.051 and for the 24-hour model from 0.262 to 0.408. The five circuits had four different busy hours. Two of the circuits had busy hours beginning at 1 a.m., one at 13:00 or 1 p.m., one at 18:00 or 6 p.m., and one at 20:00 or 8 p.m.

<table>
<thead>
<tr>
<th>ATM-HER Circuit</th>
<th>TCBH Normalized Transmission $t$</th>
<th>Busy Hour</th>
<th>24-Hour Normalized Transmission $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 in</td>
<td>0.051</td>
<td>1</td>
<td>0.408</td>
</tr>
<tr>
<td>2 in</td>
<td>0.016</td>
<td>1</td>
<td>0.404</td>
</tr>
<tr>
<td>3 in</td>
<td>0.021</td>
<td>18</td>
<td>0.348</td>
</tr>
<tr>
<td>4 in</td>
<td>0.037</td>
<td>13</td>
<td>0.262</td>
</tr>
<tr>
<td>5 in</td>
<td>0.008</td>
<td>20</td>
<td>0.269</td>
</tr>
</tbody>
</table>

Circuit usage in megabits per second (Mb/s) as a function of time shows a distinct pattern with daily and weekly cycles (Figs. 11 and 12). Similar charts for all circuits in the study are included in Appendix D.
Fig. 11. ATM circuit 1 usage over time.

Fig. 12. ATM circuit 2 usage over time.
The same data when plotted as a scatter chart of usage versus hour of the day shows a pattern of low usage during early and late morning hours with higher usage from early afternoon to late night. There was little separation between workdays and weekend usage patterns (Figs. 13 and 14). Charts for all the circuits are included in Appendix D.

Fig. 13. ATM circuit 1 usage by hour of the day.
The data from the five circuits was combined on an hour-by-hour basis for each of the 10 possible combinations (Fig. 15). One of the 10 combinations included circuits with the same busy hour. In that case there was no perceptible difference in the accuracy of the two models in predicting the combined busy hour usage. For the 9 combinations of trunks with different busy hours, the 24-hour model was more accurate than the TCBH model by 1.0% to 14.2% (Table 5).

The combination usage for Circuits 1 plus 2 is plotted over time (Fig. 16) and plotted as a scatter chart by hour of the day (Fig. 17). The usage combination charts for all the circuit combinations in the study are included in Appendix D.
Fig. 15. Usage combination of two circuits.

Fig. 16. ATM circuit 1 plus 2 usage over time.
Fig. 17. ATM circuit 1 plus 2 usage by hour of the day.

Table 5

ATM-HER Circuit Combination Model Prediction Results

<table>
<thead>
<tr>
<th>Circuits Combined</th>
<th>Combined Busy Hour</th>
<th>TCBH Model Estimate Mb/s</th>
<th>24-Hour Model Estimate Mb/s</th>
<th>TCBH Model % Error</th>
<th>24-Hour Model % Error</th>
<th>24-Hour Model % Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 and 2</td>
<td>Yes</td>
<td>670.4</td>
<td>669.8</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>0%</td>
</tr>
<tr>
<td>1 and 3</td>
<td>No</td>
<td>443.3</td>
<td>455.1</td>
<td>2.7%</td>
<td>-0.1%</td>
<td>2.7%</td>
</tr>
<tr>
<td>1 and 4</td>
<td>No</td>
<td>439.5</td>
<td>472.2</td>
<td>7.4%</td>
<td>-0.2%</td>
<td>7.4%</td>
</tr>
<tr>
<td>1 and 5</td>
<td>No</td>
<td>465.7</td>
<td>471.1</td>
<td>1.2%</td>
<td>0.2%</td>
<td>1.0%</td>
</tr>
<tr>
<td>2 and 3</td>
<td>No</td>
<td>238.9</td>
<td>251.3</td>
<td>5.2%</td>
<td>0%</td>
<td>5.2%</td>
</tr>
<tr>
<td>2 and 4</td>
<td>No</td>
<td>235.0</td>
<td>268.4</td>
<td>14.2%</td>
<td>0%</td>
<td>14.2%</td>
</tr>
<tr>
<td>2 and 5</td>
<td>No</td>
<td>261.8</td>
<td>267.3</td>
<td>2.1%</td>
<td>0.4%</td>
<td>1.7%</td>
</tr>
<tr>
<td>3 and 4</td>
<td>No</td>
<td>53.0</td>
<td>53.7</td>
<td>1.3%</td>
<td>0%</td>
<td>1.3%</td>
</tr>
<tr>
<td>3 and 5</td>
<td>No</td>
<td>50.1</td>
<td>52.6</td>
<td>5.0%</td>
<td>-0.2%</td>
<td>5.0%</td>
</tr>
<tr>
<td>4 and 5</td>
<td>No</td>
<td>63.2</td>
<td>69.7</td>
<td>10.3%</td>
<td>-0.2%</td>
<td>10.3%</td>
</tr>
</tbody>
</table>
Peering Exchange Router – Bilateral Peering Router Circuit TCBH Model Instance Versus 24-Hour Model Instance

Similar calculations on five Internet circuits that connect peering exchange routers (PXR) to bilateral peering routers show greater transmission values for the 24-hour models (Table 6). In this study the normalized transmission parameters, $t$, for the TCBH model ranged from 0.005 to 0.042 and for the 24-hour model from 0.308 to 0.362. The five circuits had three different busy hours. Three of the circuits had busy hours beginning at 15:00 or 3 p.m., one at 14:00 or 2 p.m., and one at 16:00 or 4 p.m.

Table 6

<table>
<thead>
<tr>
<th>PXR Circuit</th>
<th>TCBH Model Normalized Transmission $t$</th>
<th>Busy Hour</th>
<th>24-Hour Model Normalized Transmission $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 in</td>
<td>0.005</td>
<td>15</td>
<td>0.308</td>
</tr>
<tr>
<td>2 in</td>
<td>0.015</td>
<td>15</td>
<td>0.314</td>
</tr>
<tr>
<td>3 in</td>
<td>0.042</td>
<td>14</td>
<td>0.362</td>
</tr>
<tr>
<td>4 in</td>
<td>0.030</td>
<td>16</td>
<td>0.321</td>
</tr>
<tr>
<td>5 in</td>
<td>0.032</td>
<td>15</td>
<td>0.338</td>
</tr>
</tbody>
</table>

Circuit usage as a function of time shows a distinct pattern with daily and weekly cycles (Figs. 18 and 19). This pattern for Circuits 1 and 2 are shown as examples, and similar charts for all circuits in the study are included in Appendix E.
Fig. 18. PXR to Bilateral Peer circuit 1 usage over time.

Fig. 19. PXR to Bilateral Peer circuit 2 usage over time.
The same data when plotted as a scatter chart of usage versus hour of the day shows a pattern of low usage during late night and early morning hours with multiple patterns of higher usage from mid-morning to late evening. The upper bands of usage points are generally during workdays with the lower band during weekends. This data for Circuits 1 and 2 are shown (Figs. 20 and 21) with charts for all the circuits included in Appendix E.

Fig. 20. PXR to Bilateral Peer circuit 1 usage by hour of the day.
The data from the five circuits was combined on an hour-by-hour basis for each of the 10 possible combinations (Fig. 22). Three of the 10 combinations included circuits with the same busy hour. In those cases there was no perceptible difference in the accuracy of the two models in predicting the combined busy hour usage. For the seven combinations of circuits with different busy hours, the 24-hour model was more accurate than the TCBH model by 0.1% to 2.6% (Table 7).

The combination usage for Circuits 1 plus 2 is plotted over time (Fig. 23) and plotted as a scatter chart by hour of the day (Fig. 24). The usage combination charts for all the circuit combinations in the study are included in Appendix E.
Fig. 22. Usage combination of two circuits.

Fig. 23. PXR to Bilateral Peer circuit 1 plus 2 usage over time.
Fig. 24. PXR to Bilateral Peer circuit 1 plus 2 usage by hour of the day.

Table 7

PXR Circuit Combination Model Prediction Results

<table>
<thead>
<tr>
<th>Circuits Combined</th>
<th>Cycles</th>
<th>Combined Busy Hour Usage Mb/s</th>
<th>TCBH Model Estimate Mb/s</th>
<th>24-Hour Model Estimate Mb/s</th>
<th>TCBH Model % Error</th>
<th>24-Hour Model % Error</th>
<th>24-Hour Model % Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 and 2</td>
<td>Yes</td>
<td>1618.6</td>
<td>1618.6</td>
<td>1618.6</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>1 and 3</td>
<td>No</td>
<td>1218.4</td>
<td>1224.9</td>
<td>1218.4</td>
<td>0.5%</td>
<td>0%</td>
<td>0.5%</td>
</tr>
<tr>
<td>1 and 4</td>
<td>No</td>
<td>1612.8</td>
<td>1614.8</td>
<td>1612.7</td>
<td>0.1%</td>
<td>0%</td>
<td>0.1%</td>
</tr>
<tr>
<td>1 and 5</td>
<td>Yes</td>
<td>950.9</td>
<td>950.9</td>
<td>950.9</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>2 and 3</td>
<td>No</td>
<td>1214.4</td>
<td>1220.9</td>
<td>1214.4</td>
<td>0.5%</td>
<td>0%</td>
<td>0.5%</td>
</tr>
<tr>
<td>2 and 4</td>
<td>No</td>
<td>1607.9</td>
<td>1610.8</td>
<td>1607.9</td>
<td>0.2%</td>
<td>0%</td>
<td>0.2%</td>
</tr>
<tr>
<td>2 and 5</td>
<td>Yes</td>
<td>946.9</td>
<td>946.9</td>
<td>946.9</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>3 and 4</td>
<td>No</td>
<td>1185.8</td>
<td>1217.1</td>
<td>1186.0</td>
<td>2.6%</td>
<td>0%</td>
<td>2.6%</td>
</tr>
<tr>
<td>3 and 5</td>
<td>No</td>
<td>552.4</td>
<td>553.2</td>
<td>552.3</td>
<td>0.1%</td>
<td>0%</td>
<td>0.1%</td>
</tr>
<tr>
<td>4 and 5</td>
<td>No</td>
<td>935.9</td>
<td>943.1</td>
<td>935.9</td>
<td>0.8%</td>
<td>0%</td>
<td>0.8%</td>
</tr>
</tbody>
</table>
ATM - HER Circuit 6-hour Busy Period Model Instance Versus TCBH and 24-Hour Model Instances

The next study involves a comparison of three different types of models. The 6-hour model has four possible busy periods: period B, beginning at midnight; period C, beginning at 6 a.m.; period D, beginning at noon; and period E, beginning at 18:00 or 6 p.m. The results from this set of models is compared to previous study results for the TCBH hour model and 24-hour model from the same five ATM-HER circuits from the previous study (Tables 8 and 9). In this study, the normalized transmission parameters, $t$, for the six consecutive hour model ranged from 0.067 to 0.253. The five circuits had two different 6-hour busy periods. Three of the circuits had busy periods beginning at 18:00 or 6 p.m., and two at noon.

Table 8

ATM-HER Circuit 6-Hour Period Mutual Information (Transmission) Results

<table>
<thead>
<tr>
<th>ATM-HER Circuit</th>
<th>Model Normalized Transmission $t$</th>
<th>Busy Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 in</td>
<td>0.067</td>
<td>E</td>
</tr>
<tr>
<td>2 in</td>
<td>0.133</td>
<td>E</td>
</tr>
<tr>
<td>3 in</td>
<td>0.217</td>
<td>D</td>
</tr>
<tr>
<td>4 in</td>
<td>0.253</td>
<td>D</td>
</tr>
<tr>
<td>5 in</td>
<td>0.086</td>
<td>E</td>
</tr>
</tbody>
</table>

The data from the five circuits was combined on an hour-by-hour basis for each of the 10 possible combinations. Four of the 10 combinations included circuits with the same busy period. The 6-hour model had a zero error in predicting the combined circuit busy period usage in five of the ten cases (Table 9).
Comparing the normalized transmission for the TCBH, 6-hour, and 24-hour models shows significant differences among most models (Table 10). The normalized transmission is the greatest for the 24-hour model and the least for the TCBH model.

Table 9
ATM-HER Circuit Combination 6-Hour Period Model Prediction Results

<table>
<thead>
<tr>
<th>Circuits Combined</th>
<th>Combined Busy Period Usage Mb/s</th>
<th>Busy Period Model Average Mb/s</th>
<th>Busy Period Model Estimate Mb/s</th>
<th>Busy Period Model % Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 and 2</td>
<td>Yes 586.0</td>
<td>586.0</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>1 and 3</td>
<td>No 391.0</td>
<td>391.1</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>1 and 4</td>
<td>No 393.0</td>
<td>390.3</td>
<td>-0.7%</td>
<td></td>
</tr>
<tr>
<td>1 and 5</td>
<td>Yes 408.8</td>
<td>409.6</td>
<td>0.2%</td>
<td></td>
</tr>
<tr>
<td>2 and 3</td>
<td>No 219.9</td>
<td>220.0</td>
<td>0.1%</td>
<td></td>
</tr>
<tr>
<td>2 and 4</td>
<td>No 221.9</td>
<td>219.2</td>
<td>-1.2%</td>
<td></td>
</tr>
<tr>
<td>2 and 5</td>
<td>Yes 237.8</td>
<td>238.6</td>
<td>0.3%</td>
<td></td>
</tr>
<tr>
<td>3 and 4</td>
<td>Yes 48.4</td>
<td>48.4</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>3 and 5</td>
<td>No 40.7</td>
<td>40.7</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>4 and 5</td>
<td>No 56.5</td>
<td>56.5</td>
<td>0.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 10
ATM-HER Combination Circuit Mutual Information (Transmission) Results

<table>
<thead>
<tr>
<th>ATM-HER Circuit</th>
<th>TCBH Normalized Transmission t</th>
<th>6-Hour Model Normalized Transmission t</th>
<th>24-Hour Model Normalized Transmission t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.051</td>
<td>0.067</td>
<td>0.408</td>
</tr>
<tr>
<td>2</td>
<td>0.016</td>
<td>0.133</td>
<td>0.404</td>
</tr>
<tr>
<td>3</td>
<td>0.021</td>
<td>0.217</td>
<td>0.348</td>
</tr>
<tr>
<td>4</td>
<td>0.037</td>
<td>0.253</td>
<td>0.262</td>
</tr>
<tr>
<td>5</td>
<td>0.008</td>
<td>0.086</td>
<td>0.269</td>
</tr>
<tr>
<td>Average</td>
<td>0.027</td>
<td>0.151</td>
<td>0.338</td>
</tr>
</tbody>
</table>
Error for each model was defined as the difference between the busy hour or busy period usage of the combination circuits and the busy hour or busy period of the combination of the model usage in the busy hour or busy period. The Circuit 1 error was determined by adding the absolute value of the error of each combination of circuits that included Circuit 1. The error results for the other four circuits were calculated in a similar manner. The results (Table 11) show that the 24-hour model is the most accurate and the TCBH model is the least accurate.

Table 11

<table>
<thead>
<tr>
<th>ATM-HER Circuit Combination</th>
<th>TCBH Model Error</th>
<th>6-Hour Model Error</th>
<th>24-Hour Model Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuit 1</td>
<td>11.4%</td>
<td>0.91%</td>
<td>0.60%</td>
</tr>
<tr>
<td>Circuit 2</td>
<td>21.6%</td>
<td>1.59%</td>
<td>0.50%</td>
</tr>
<tr>
<td>Circuit 3</td>
<td>14.2%</td>
<td>0.09%</td>
<td>0.30%</td>
</tr>
<tr>
<td>Circuit 4</td>
<td>33.2%</td>
<td>1.89%</td>
<td>0.40%</td>
</tr>
<tr>
<td>Circuit 5</td>
<td>18.6%</td>
<td>0.53%</td>
<td>1.00%</td>
</tr>
<tr>
<td>Average</td>
<td>19.8%</td>
<td>1.00%</td>
<td>0.56%</td>
</tr>
</tbody>
</table>

Summary

In this chapter, the results of modeling traffic usage data for telephony voice communications trunks and Internet usage data for two types of circuits are described. TCBH models are compared to 24-hour models for each case. For one of the Internet circuit types, a third model with a 6-hour busy period is included.
CHAPTER 6
DISCUSSION

Complex Communications Network Modeling and Decomposition by Entropic Techniques

The results of this study support the original hypothesis. The normalized transmission parameter was a useful predictor of the relative accuracy of a model in predicting the effects of combining usage on trunks or circuits where there was a significant difference in the model parameters and the trunks or circuits had dissimilar busy hours or busy periods. Most models performed with similar accuracy when the trunks or circuits to be combined had identical busy hours or busy periods. When multiple models were compared, the average accuracy was consistent with the normalized transmission parameter.

The models used in this study are simple in order to illustrate the process used and so that graphical plots can be used intuitively to visually confirm the results. Previous studies focused on the methods required to compute transmission parameters but did not provide a method to confirm their approach. Using parameters from combinations of data sets is a useful approach since network traffic from multiple sources is often combined in common circuits or devices.

For any model to be useful in the sizing and timing of capacity additions there have to be corresponding service-level objectives that are relevant to the model. For example, an objective of no more than 2% blocking during a single busy hour period
would require an adjustment to provide the same effective service level using a 6-hour busy period. Other models that have been used to simplify system programming include a “bouncing” busy hour. In this model, the peak usage hour in each workday is used instead of a TCBH. This model tends to trigger more capacity than a time-consistent model for the same usage and service-level objectives.

Models used for real-time capacity monitoring often focus on usage over intervals much shorter than an hour and often use peak values or threshold crossings when comparing measurements to service objectives. These process models are used to trigger immediate corrective action to respond to equipment or software failures, unusual traffic loads, or denial of service attacks. The data used for these purposes is generally discarded within weeks to minimize the cost of storage and therefore is less useful for identifying long term or cyclic trends.

In any study, whether academic or made for economic decision making, the accuracy of the underlying data is essential. Efforts to identify and accommodate missing or invalid data points can easily exceed the effort required for the study itself. Approaches such as entropic measures are less sensitive to a few erroneous data points than the trending or averaging of data. Future studies could be used to confirm this hypothesis.

An Exploration of Other Applications of Decomposition by Entropic Techniques

*Service-Based Entropic Model for Sensors Orchestration*

A service-based approach to decomposition and orchestration for managing a self-configuring cluster network of sensors is described by Sadasivam, Goli, Kathiru,
Krishnan, Thompson, Tuncer, and Tanik [29]. The sensors are used by the U.S. Navy to
detect and track surface ships and submarines. This approach uses the technique of
decomposition by transmission to identify clusters of busy sensors that can be
reconfigured to conserve power and reduce the probability of detection.

Mapping Networks of Terrorist Cells

The difficulty of mapping covert networks is the subject of a paper by Krebs [30].
This topic has seen increased interest since the terrorist events of September 11, 2001.
Krebs creates relationship network maps using publicly available information to illustrate
the structure of terrorist networks. This mapping after the event is a helpful tool to the
prevention of future attacks.

While not described by Krebs, this approach could possibly be enhanced through
the use of decomposition by transmission. The same data sources used by Krebs could be
used to identify all individuals who share some common data point with known terrorists.
The multiple sets of data could then be considered multiple observations for each
individual in each list. The decomposition by transfer could then be used to identify
individuals from the separate data sources who have some relationship stronger than
chance.

Summary

The results of this study support the original hypothesis. The normalized
transmission parameter was a useful predictor of the relative accuracy of a model in
predicting the effects of combining usage on trunks or circuits where there was a
significant difference in the model parameters and the trunks or circuits had dissimilar busy hours or busy periods. The models used in this study are simple in order to illustrate the process used and so that graphical plots can be used intuitively to visually confirm the results. Previous studies focused on the methods required to compute transmission parameters but did not provide a method to confirm their approach. The purpose of the model is a key factor in the selection or design of an appropriate model.

There are other applications for decomposition by transmission and for models that use this approach. Example applications in the design of sensor networks and in the mapping of social networks have been described to illustrate the wide variety of potential applications.
CHAPTER 7

CONCLUSION

Elements of a complex system may have measures that can be used to describe their internal processes as well as their interactions with other system elements. This work has shown how the information content in these measures can provide a signature that can be used to compare the activity of the element to a simplified model of the system. The simplified models can then be used in engineering, typically to design timely, reliable, and economical architectures for changing complex systems.

The issues related to the construction of appropriate measures to use when calculating the entropy of a set of observations of a study variable have been examined.

The application of Conant’s transmission calculations to the decomposition of complex systems have been described and extended with a discussion of the precision versus accuracy versus validity of study data.

Procedures for performing necessary calculations have been developed and described in detail. Mechanized programs have been written that apply these procedures to samples of telephony and Internet usage data sets.

The manner in which information theory can be used to improve the economic design of data and voice network architectures and how entropic measures can be used to evaluate simplified models of Internet circuit usage patterns has been described.

Other applications of decomposition by entropic techniques have been explored, such as a tool for configuring sensor networks to save energy and reduce the probability
of their detection. Also, an approach to the identification of individuals with relationships to terrorist networks by using decomposition by transmission techniques has been described.
LIST OF REFERENCES


APPENDIX A

JAVA PROGRAMS FOR CALCULATIONS
This section contains pseudocode listings for the Java programs used in this study:

**TestObservationDataSet.java** – A program to process data sets and compute the transmission parameter for the actual data compared to a time consistent model.

Pseudocode:

For each command line argument

Create an instance of an ObservationDataSet using the argument as the filename.

Print the filename to the screen

**ObservationDataSet.java** – This class defines the parameters of a set of time series observations.

Pseudocode:

Construct an instance of class ObservationDataSet.

Count the number of records in the data set file using method countRecords().

Print the file name and number of records to the screen.

Create data arrays using method readRecords().

Count the number of workday records using method setrNumberOfWorkdays().

Print the number of workday records to the screen.

Compute the average usage for each workday hour using method computeTCBH().

Print the time-consistent busy-hour using method displayTCBH().
Display the average usage for each workday hour using method 
\texttt{displayWorkdayUsageAverage()}. 
Create an array of 24 threshold values based on workday usage using method 
\texttt{createThresholdArray()}. 
Create and sort a String array from the thresholds and all usage records using 
method \texttt{createSortedLabelArray()}. 
Count the number of each unique String label in the string array including 
workdays, weekends and holidays using method \texttt{createArrayOfLabelCounts()}. 
Compute the entropy of the set of 24 labels using static method 
\texttt{entropyCalculator()} from class \texttt{TransCalc}. 
Print the entropy to the screen. 
Compute the entropy of all usage value labels in the data set using static method 
\texttt{entropyCalculator()} from class \texttt{TransCalc}. 
Print the entropy to the screen. 
Create and sort a String array of combinations of model and observation set labels 
using methods 
\texttt{createSortedWorkdayUsageAverageArray()},\texttt{createArrayOfComboLabels()}, and 
by creating an instance of class \texttt{SortByMerge}. 
Count the number of each unique combination of String labels and compute the 
entropy using method \texttt{createArrayOfLabelCounts()} and static method 
\texttt{entropyCalculator()} from class \texttt{TransCalc}. 
Print the entropy to the screen.
Compute transmission and normalized transmission using static methods computeTransmission() and computeNormTrans() from class TransCalc.
Print the transmission and normalized transmission to the screen.
Count the number of each unique combinations of String labels and compute the entropy of the TCBH model using methods createCountTCBHComboArray() and createCountTCBHModelArray() and static method entropyCalculator from class TransCalc.
Print entropy values to the screen.
Compute transmission and normalized transmission of TCBH model.

Method computeTCBH:
Create an array of workday usage records.
For every observation record.
   Check to see if it is a workday record.
   If true add a workday usage record = observation usage.
Count and sum workday usage records by hour For hours 0 to 23
If observation hour = hour
   Increment the hour count
   Add the observation usage to the workday usage sum
Check for a maximum usage value using static method
max() of class Math.

For each hour, 0 to 23, compute average usage and TCBH average usage.

Workday usage average = workday usage sum / hour count.

Check for busy hour by comparing each hour’s usage to the maximum using static method max() from class Math.

Method countRecords:

Open the file and read each line.
Increment a counter until there are no more records.

Method createArrayOfComboLabels:
For each observation record

Concatenate the label array value with the observation hour

Method createArrayOfLabelCounts:
Count the total number of unique labels in a sorted array of labels
Create an array of label counts using the number of labels as the size of the array.
There is at least one label.
Since the array to count has been sorted, you count each label that matches the first label and put that total in the first member of the array of label counts.
When you reach a label that does not match the preceding label, begin a new count.

Method createCountTCBHModelArray;
For observation record
  If the observation usage is greater than the 23rd threshold value
    Increment count of TCBH model array 0
  If the observation usage is less than or equal to the 22nd threshold value
    Increment count of TCBH model array 1

Method createSortedLabelArray:
  Compare each observation usage value to the threshold array and create a label in the unsorted label array.
  Sort the label array by creating an instance of the SortByMerge class.

Method createSortedWorkdayUsageAverageArray:
  Sort the workday usage average array by creating an instance of the SortByMerge class.

Method createThresholdArray:
  Sort the workday usage array by creating an instance of the SortByMerge class.
  For each hour 0 to 23
Compute the threshold span by dividing the number of workday records by 24 and multiplying by 1 plus the hour.

TransCalc.java – This class contains methods to calculate log to the base two, entropy, transmission and normalized transmission.

Pseudocode:

Method computeTransmission:

Transmission = model entropy plus vector entropy – combo entropy.

Method computeNormTrans:

Normalized transmission = transmission / vector entropy.

Method getLogTwo:

If input is greater than zero.

Use static method log in class Math to compute log(input)/log(2).

Method counterLogCounter:

If input is greater than zero.

Use method getLogTwo to return input times log₂(input).
Method entropyCalculator:

The input is an array of integers representing the counts of each state in a set of observations.
Compute the number of observations, \( N \), by summing the array values.
Use method counterLogCounter to compute the sum of the input times \( \log_2(\text{input}) \)
and sum for all array input values.
Entropy of the array = \( \log_2(N) – \frac{1}{N} \) times the sum of the input times \( \log_2(\text{input}) \)

SortByMerge.java – This class contains methods for sorting arrays.
Pseudocode:
Construct an instance of class SortByMerge.
Use method sort for double or float type data.
Use method stringSort for String type data.

Method sort:
Use method sortArray to sort an input array.

Method stringSort:
Use method stringSortArray to sort an input array.

Method sortArray:
Test for base case where size of array equals one element.

If the largest index minus the smallest index is greater than zero then this is not the base case.

Calculate the middle of array = (largest index plus smallest index) / 2.

Middle plus one is the low index for the upper split of the split array.

Split array in half; sort each half (recursive calls to method sortArray).

Merge sorted arrays using method merge.

Method stringSortArray:

Test for base case where size of array equals one element.

If the largest index minus the smallest index is greater than zero then this is not the base case.

Calculate the middle of array = (largest index plus smallest index) / 2.

Middle plus one is the low index for the upper split of the split array.

Split array in half; sort each half (recursive calls to method sortArray).

Merge sorted arrays using method stringMerge.

Method merge:

Merge arrays until reaching end of either.

While the index of the left array is less than or equal to the middle index and the index of the right array is less than the largest index
Place smaller of two current elements into result and move to next space in arrays.

If left array is empty

Copy in rest of right array.

Else right array is empty

Copy in rest of left array.

Method stringMerge:

Merge arrays until reaching end of either.

While the index of the left array is less than or equal to the middle index and the index of the right array is less than the largest index

Place smaller of two current elements into result and move to next space in arrays.

If left array is empty

Copy in rest of right array.

Else right array is empty

Copy in rest of left array.

ObservationRecord.java – This class defines the variables in an observation record.

Pseudocode:

Construct an instance of class ObservationRecord:
Use methods setObservationNumber, setObservationDate, setObservationDay,
setObservationPeriod, setObservationAmPm, setObservationHour,
setObservationUsage, setObservationValidDate, getObservationDateMonth,
getObservationDateDay, getObservationYear, setObservationValidTime,
setObservationHoliday.

Method setObservationNumber:

The observation number equals the input.

If the input is less than the previous input

Print a message.

If the input is greater than the previous input by more than one

Print a message.

Method setObservationDate:

Observation date equals input.

Method setObservationDay:

Observation day equals input.

Method setObservationPeriod:

Observation period equals input.
Method setObservationAmPm:

Observation equals input.

Method setObservationHour:

Observation hour equals input.

Method setObservationUsage:

If input is greater than 1

Observation usage equals input.

If input is less than 1

Observation equals 1.

Method setObservationValidDate:

Construct an instance of class Date.

Method getObservationDateMonth:

Observation date month equals 0 until set by method setObservationDateMonth.

Method getObservationDateDay:

Observation date day equals 0 until set by method setObservationDateDay.

Method getObservationYear:

Observation year equals 0 until set by method setObservationYear.
Method setObservationValidTime:

Create an instance of class TimeValidator.

Method setObservationHoliday:

Set observation holiday flag to false.

If observation date month is January and observation date day is 1
    Set observation holiday flag to true.
If observation date month is December and observation date day is 31 and
observation day is Friday
    Set observation holiday flag to true.
If observation date month is January and observation date day is 2 and
observation day is Monday
    Set observation holiday flag to true.
If observation date month is May and observation date day is greater than 24 and
observation day is Monday
    Set observation holiday flag to true.
If observation date month is July and observation date day is 4
    Set observation holiday flag to true.
If observation date month is July and observation date day is 3 and observation
day is Friday
Set observation holiday flag to true.

If observation date month is July and observation date day is 5 and observation day is Monday
    Set observation holiday flag to true.

If observation date month is September and observation date day is less than 8 and observation day is Monday
    Set observation holiday flag to true.

If observation date month is November and observation date day is greater than 21 and less than 29 and observation day is Thursday
    Set observation holiday flag to true.

If observation date month is December and observation date day is 25
    Set observation holiday flag to true.
If observation date month is December and observation date day is 24 and observation day is Friday
    Set observation holiday flag to true.
If observation date month is December and observation date day is 26 and observation day is Monday
    Set observation holiday flag to true.
Method setObservationDateMonth:

Observation date month equals the first two digits of the input date.

If the previous observation date month is not equal to 12 and observation date
month is less than previous date month

Print a message.

Method setObservationDateDay:

Observation date day equals the fourth and fifth digits of the input date.

If observation date day is not equal to 1 and is less than the previous observation
date day

Print a message.

If previous observation period hour is not equal to 24 and the observation date day
is not equal to the previous observation date day and the observation number is
not equal to 1

Print a message.

Method setObservationYear:

If the length of the input date string is less than ten characters

Observation year equals 2003.

Else the observation year equals the seventh through the tenth characters of the
input date string.
TimeValidator.java – This class is used to ensure that data records contain valid time values.

Pseudocode:

Construct an instance using method setTime.

Method setTime:

Set the hour using method setHour.
Set the minute using method setMinute.
Set the second using method setSecond.

Method setHour:

If input is greater than or equal to 0 and less than 24
    Hour equals input.
Else hour equals 0.

Method setMinute:

If input is greater than or equal to 0 and less than 60
    Minute equals input.
Else minute equals 0.

Method setSecond:

If input is greater than or equal to 0 and less than 60
    Second equals input.
Else second equals 0.

Date.java – This class is used to ensure that data records contain valid time values.

Pseudocode:

Construct an instance of a Date class

    Use checkMonth method to validate the month
    Use checkYear method to validate the year
    Use checkDay method to validate the day

Print the date to the screen

Method checkYear:

    If the year is between 1948 and 2050 return the year
    Else return the year as 1 and print an error message to the screen

Method checkMonth:

    If the month is greater than 0 and less than or equal to 12, return the month
    Else return the month as 1 and print an error message to the screen

Method checkDay:

    Create an array consisting of the number of days in each calendar month
    If the day is greater than 0 and less than or equal to the number of days in the
    month, return the day
If the year is a leap year and the month is 2 and the day is 29, return the day

Else return the day as 1 and print an error message to the screen
APPENDIX B

SAMPLE SPREADSHEET FOR 6-HOUR MODELS
This section contains sample spreadsheet listings showing formulas for calculating the 6-hour model parameters. Redundant rows are omitted.
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Circuit</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>High Q</td>
<td>Average Busy Quartile</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.0859414641909568</td>
<td>0.1482899162797953</td>
<td>0.0965208054047736</td>
<td>0.0669907793214766</td>
<td>0.355268115809354</td>
<td>415.25327510917</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.0428206188887192</td>
<td>0.147632532521435</td>
<td>0.1265976297535</td>
<td>0.133135208912233</td>
<td>0.361937418746801</td>
<td>224.222222222222</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0.144906242076577</td>
<td>0.145221383921123</td>
<td>0.217137678134436</td>
<td>0.0146435145081641</td>
<td>0.316053087013601</td>
<td>17.133759617834</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0.145137605931938</td>
<td>0.0568169245559764</td>
<td>0.25292296624371</td>
<td>7.90312917664885E-00</td>
<td>0.258600571763929</td>
<td>31.6</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>0.0035066497628941</td>
<td>0.1289703153335992</td>
<td>0.00778935668814533</td>
<td>0.0954824787140354</td>
<td>34.192657428571</td>
<td>31.2700421940928</td>
</tr>
<tr>
<td>8</td>
<td>1 plus 2</td>
<td>0.0669750609450291</td>
<td>0.149980428982736</td>
<td>0.1146678814592226</td>
<td>0.046888213234537</td>
<td>0.347738489076763</td>
<td>637.701754385965</td>
</tr>
<tr>
<td>9</td>
<td>1 plus 3</td>
<td>0.0767175287349933</td>
<td>0.1499659780906822</td>
<td>0.0929604975806084</td>
<td>0.0748903760679957</td>
<td>0.342211617212119</td>
<td>423.76</td>
</tr>
<tr>
<td>10</td>
<td>1 plus 4</td>
<td>0.069120263338444</td>
<td>0.149171575909357</td>
<td>0.087373291305915</td>
<td>0.0787213493680195</td>
<td>0.331602526508945</td>
<td>420.415178571429</td>
</tr>
<tr>
<td>11</td>
<td>1 plus 5</td>
<td>0.0719067984179677</td>
<td>0.149171575909357</td>
<td>0.0922731807346669</td>
<td>0.0787213493680195</td>
<td>0.34672401662233</td>
<td>444.955357142857</td>
</tr>
<tr>
<td>12</td>
<td>2 plus 3</td>
<td>0.031002102719169</td>
<td>0.14522126836665</td>
<td>0.0939676718939832</td>
<td>0.137732625078474</td>
<td>0.357369389178471</td>
<td>233.602739726027</td>
</tr>
<tr>
<td>13</td>
<td>2 plus 4</td>
<td>0.0093517257809292</td>
<td>0.147587098302257</td>
<td>0.025507540121902</td>
<td>0.116903370561541</td>
<td>0.201036112791491</td>
<td>232.387387387387</td>
</tr>
<tr>
<td>14</td>
<td>2 plus 5</td>
<td>0.0397518570235197</td>
<td>0.14679701390687</td>
<td>0.106590677600174</td>
<td>0.12927140150762</td>
<td>0.365836728357403</td>
<td>254.701357466063</td>
</tr>
<tr>
<td>15</td>
<td>3 plus 4</td>
<td>0.145137605931938</td>
<td>0.0813795924363461</td>
<td>0.248773607063141</td>
<td>0.0012814254557601</td>
<td>0.26643642143958</td>
<td>47.8857142857143</td>
</tr>
<tr>
<td>16</td>
<td>3 plus 5</td>
<td>0.042815637160231</td>
<td>0.139233047644648</td>
<td>0.032859523820896</td>
<td>0.0828817793820096</td>
<td>0.19879118145708</td>
<td>47.5061056105611</td>
</tr>
<tr>
<td>17</td>
<td>4 plus 5</td>
<td>0.12637015831515</td>
<td>0.139103764093551</td>
<td>0.24097689818129</td>
<td>0.0073588641753450</td>
<td>0.290353884346558</td>
<td>57.8930817610063</td>
</tr>
</tbody>
</table>

Normalized Transmission Usage
<table>
<thead>
<tr>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average Busy Quartile</td>
<td>Average Six High Busy Hours</td>
<td>Sum of Circuits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>366.33152173913</td>
<td>157.091954022989</td>
<td>292.98275862069</td>
<td>378.520231213873</td>
<td>=G3+G4</td>
<td>=H3+H4</td>
<td>=I3+I4</td>
</tr>
<tr>
<td>4</td>
<td>192.620111731844</td>
<td>77.958333333333</td>
<td>161.035714285714</td>
<td>207.456647398844</td>
<td>=G3+G5</td>
<td>=H3+H5</td>
<td>=I3+I5</td>
</tr>
<tr>
<td>5</td>
<td>5.42857142857143</td>
<td>5.38626609442609</td>
<td>16.2649572649572</td>
<td>12.5815895861586</td>
<td>=G3+G6</td>
<td>=H3+H6</td>
<td>=I3+I6</td>
</tr>
<tr>
<td>6</td>
<td>1.78947368421053</td>
<td>4.24757281553398</td>
<td>32.1324786324786</td>
<td>11.7647058823529</td>
<td>=G3+G7</td>
<td>=H3+H7</td>
<td>=I3+I7</td>
</tr>
<tr>
<td>7</td>
<td>26.0548523020675</td>
<td>10.6094420600859</td>
<td>24.393623931624</td>
<td>31.1087866108787</td>
<td>=G3+G8</td>
<td>=H3+H8</td>
<td>=I3+I8</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>560.050279329609</td>
<td>235.422619047619</td>
<td>455.130952380952</td>
<td>585.976878612717</td>
<td>=G5+G4</td>
<td>=H5+H4</td>
</tr>
<tr>
<td>9</td>
<td>1372.748603351955</td>
<td>162.678571428571</td>
<td>310.089285714286</td>
<td>390.971098265896</td>
<td>=G5+G5</td>
<td>=H5+H5</td>
<td>=I5+I5</td>
</tr>
<tr>
<td>10</td>
<td>369.117318435754</td>
<td>162.184523809524</td>
<td>333.607142857143</td>
<td>392.596375722554</td>
<td>=G5+G6</td>
<td>=H5+H6</td>
<td>=I5+I6</td>
</tr>
<tr>
<td>11</td>
<td>390.301675977654</td>
<td>167.75</td>
<td>317.285714285714</td>
<td>408.826895959595</td>
<td>=G5+G7</td>
<td>=H5+H7</td>
<td>=I5+I7</td>
</tr>
<tr>
<td>12</td>
<td>197.938547496034</td>
<td>83.1726190476191</td>
<td>177.029761904762</td>
<td>219.907514450867</td>
<td>=G5+G8</td>
<td>=H5+H8</td>
<td>=I5+I8</td>
</tr>
<tr>
<td>13</td>
<td>194.307262569832</td>
<td>82.6785714285714</td>
<td>200.611904761904</td>
<td>221.895952380952</td>
<td>=G6+G6</td>
<td>=H6+H6</td>
<td>=I6+I6</td>
</tr>
<tr>
<td>14</td>
<td>217.061452513966</td>
<td>87.5</td>
<td>183.214285714286</td>
<td>237.780346820809</td>
<td>=G6+G7</td>
<td>=H6+H7</td>
<td>=I6+I7</td>
</tr>
<tr>
<td>16</td>
<td>31.4894514767932</td>
<td>15.9597081545064</td>
<td>40.6581196581971</td>
<td>43.6903765690377</td>
<td>=G7+G6</td>
<td>=H7+H6</td>
<td>=I7+I6</td>
</tr>
<tr>
<td>17</td>
<td>27.6329113924051</td>
<td>14.3648086869528</td>
<td>56.525646029564</td>
<td>42.824267824268</td>
<td>=G7+G7</td>
<td>=H7+H7</td>
<td>=I7+I7</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>R</td>
<td>S</td>
<td>T</td>
<td>U</td>
<td>V</td>
<td>W</td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>=K3+K4 =L3+L4 =M10/G10-1 =N10/H10-1 =O10/I10-1 =P10/J10-1 =Q10/K10-1 =R10/L10-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>=K3+K6 =L3+L6 =M12/G12-1 =N12/H12-1 =O12/I12-1 =P12/J12-1 =Q12/K12-1 =R12/L12-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>=K3+K7 =L3+L7 =M13/G13-1 =N13/H13-1 =O13/I13-1 =P13/J13-1 =Q13/K13-1 =R13/L13-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>=K4+K5 =L4+L5 =M14/G14-1 =N14/H14-1 =O14/I14-1 =P14/J14-1 =Q14/K14-1 =R14/L14-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>=K4+K6 =L4+L6 =M15/G15-1 =N15/H15-1 =O15/I15-1 =P15/J15-1 =Q15/K15-1 =R15/L15-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>=K4+K7 =L4+L7 =M16/G16-1 =N16/H16-1 =O16/I16-1 =P16/J16-1 =Q16/K16-1 =R16/L16-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>=K5+K6 =L5+L6 =M17/G17-1 =N17/H17-1 =O17/I17-1 =P17/J17-1 =Q17/K17-1 =R17/L17-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>=K5+K7 =L5+L7 =M18/G18-1 =N18/H18-1 =O18/I18-1 =P18/J18-1 =Q18/K18-1 =R18/L18-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>=K6+K7 =L6+L7 =M19/G19-1 =N19/H19-1 =O19/I19-1 =P19/J19-1 =Q19/K19-1 =R19/L19-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This appendix contains charts for each file used in the study. There are two sets of charts. The first is a scatter plot of the observation usage versus the observation hour. This format displays the time of day usage patterns. The second set of charts show usage over time. This format displays the daily and weekly patterns. Some of the data sets are combinations of other sets. A diagram is included at the beginning of the charts to illustrate how the data sets are combined.

Fig. 25. Usage combination of two trunk groups.

Fig. 26. Trunk 1 usage by hour of the day.

Fig. 27. Trunk 2 usage by hour of the day.

Fig. 28. Trunk 3 usage by hour of the day.
Fig. 29. Trunk 4 usage by hour of the day.

Fig. 30. Trunk 5 usage by hour of the day.

Fig. 31. Trunk 1 usage over time.

Fig. 32. Trunk 2 usage over time.

Fig. 33. Trunk 3 usage over time.

Fig. 34. Trunk 4 usage over time.
Fig. 35. Trunk 5 usage over time.

Fig. 36. Trunk 1 plus 2 usage by hour of the day.

Fig. 37. Trunk 1 plus 3 usage by hour of the day.

Fig. 38. Trunk 1 plus 4 usage by hour of the day.

Fig. 39. Trunk 1 plus 5 usage by hour of the day.

Fig. 40. Trunk 2 plus 3 usage by hour of the day.
Fig. 41. Trunk 2 plus 4 usage by hour of the day.

Fig. 42. Trunk 2 plus 5 usage by hour of the day.

Fig. 43. Trunk 3 plus 4 usage by hour of the day.

Fig. 44. Trunk 3 plus 5 usage by hour of the day.

Fig. 45. Trunk 4 plus 5 usage by hour of the day.

Fig. 46. Trunk 1 plus 2 usage over time.
Fig. 47. Trunk 1 plus 3 usage over time.

Fig. 48. Trunk 1 plus 4 usage over time.

Fig. 49. Trunk 1 plus 5 usage over time.

Fig. 50. Trunk 2 plus 3 usage over time.

Fig. 51. Trunk 2 plus 4 usage over time.

Fig. 52. Trunk 2 plus 5 usage over time.
Fig. 53. Trunk 3 plus 4 usage over time.

Fig. 54. Trunk 3 plus 5 usage over time.

Fig. 55. Trunk 4 plus 5 usage over time.
APPENDIX D

SAMPLE ATM-HER DATA
This appendix contains charts for each file used in the study. There are two sets of charts. The first is a scatter plot of the observation usage versus the observation hour. This format displays the time of day usage patterns. The second set of charts show usage over time. This format displays the daily and weekly patterns. Some of the data sets are combinations of other sets. A diagram is included at the beginning of the charts to illustrate how the data sets are combined.

Fig. 56. Usage combination of two circuits.

Fig. 57. ATM circuit 1 usage by hour of the day.

Fig. 58. ATM circuit 2 usage by hour of the day.

Fig. 59. ATM circuit 3 usage by hour of the day.
Fig. 60. ATM circuit 4 usage by hour of the day.

Fig. 61. ATM circuit 5 usage by hour of the day.

Fig. 62. ATM circuit 1 usage over time.

Fig. 63. ATM circuit 2 usage over time.

Fig. 64. ATM circuit 3 usage over time.

Fig. 65. ATM circuit 4 usage over time.
Fig. 66. ATM circuit 5 usage over time.

Fig. 67. ATM circuit 1 plus 2 usage by hour of the day.

Fig. 68. ATM circuit 1 plus 3 usage by hour of the day.

Fig. 69. ATM circuit 1 plus 4 usage by hour of the day.

Fig. 70. ATM circuit 1 plus 5 usage by hour of the day.

Fig. 71. ATM circuit 2 plus 3 usage by hour of the day.
Fig. 72. ATM circuit 2 plus 4 usage by hour of the day.

Fig. 73. ATM circuit 2 plus 5 usage by hour of the day.

Fig. 74. ATM circuit 3 plus 4 usage by hour of the day.

Fig. 75. ATM circuit 3 plus 5 usage by hour of the day.

Fig. 76. ATM circuit 4 plus 5 usage by hour of the day.

Fig. 77. ATM circuit 1 plus 2 usage over time.
Fig. 78. ATM circuit 1 plus 3 usage over time.

Fig. 79. ATM circuit 1 plus 4 usage over time.

Fig. 80. ATM circuit 1 plus 5 usage over time.

Fig. 81. ATM circuit 2 plus 3 usage over time.

Fig. 82. ATM circuit 2 plus 4 usage over time.

Fig. 83. ATM circuit 2 plus 5 usage over time.
Fig. 84. ATM circuit 3 plus 4 usage over time.

Fig. 85. ATM circuit 3 plus 5 usage over time.

Fig. 86. ATM circuit 4 plus 5 usage over time.
APPENDIX E

SAMPLE PXR-BILATERAL PEER DATA
This appendix contains charts for each file used in the study. There are two sets of charts. The first is a scatter plot of the observation usage versus the observation hour. This format displays the time of day usage patterns. The second set of charts show usage over time. This format displays the daily and weekly patterns. Some of the data sets are combinations of other sets. A diagram is included at the beginning of the charts to illustrate how the data sets are combined.

Fig. 87. Usage combination of two circuits.

Fig. 88. PXR to Bilateral Peer circuit 1 usage by hour of the day.

Fig. 89. PXR to Bilateral Peer circuit 2 usage by hour of the day.

Fig. 90. PXR to Bilateral Peer circuit 3 usage by hour of the day.
Fig. 91. PXR to Bilateral Peer circuit 4 usage by hour of the day.

Fig. 92. PXR to Bilateral Peer circuit 5 usage by hour of the day.

Fig. 93. PXR to Bilateral Peer circuit 1 usage over time.

Fig. 94. PXR to Bilateral Peer circuit 2 usage over time.

Fig. 95. PXR to Bilateral Peer circuit 3 usage over time.

Fig. 96. PXR to Bilateral Peer circuit 4 usage over time.
Fig. 97. PXR to Bilateral Peer circuit 5 usage over time.

Fig. 98. PXR to Bilateral Peer circuit 1 plus 2 usage by hour of the day.

Fig. 99. PXR to Bilateral Peer circuit 1 plus 3 usage by hour of the day.

Fig. 100. PXR to Bilateral Peer circuit 1 plus 4 usage by hour of the day.

Fig. 101. PXR to Bilateral Peer circuit 1 plus 5 usage by hour of the day.

Fig. 102. PXR to Bilateral Peer circuit 2 plus 3 usage by hour of the day.
Fig. 103. PXR to Bilateral Peer circuit 2 plus 4 usage by hour of the day.

Fig. 104. PXR to Bilateral Peer circuit 2 plus 5 usage by hour of the day.

Fig. 105. PXR to Bilateral Peer circuit 3 plus 4 usage by hour of the day.

Fig. 106. PXR to Bilateral Peer circuit 3 plus 5 usage by hour of the day.

Fig. 107. PXR to Bilateral Peer circuit 4 plus 5 usage by hour of the day.

Fig. 108. PXR to Bilateral Peer circuit 1 plus 2 usage over time.
Fig. 109. PXR to Bilateral Peer circuit 1 plus 3 usage over time.

Fig. 110. PXR to Bilateral Peer circuit 1 plus 4 usage over time.

Fig. 111. PXR to Bilateral Peer circuit 1 plus 5 usage over time.

Fig. 112. PXR to Bilateral Peer circuit 2 plus 3 usage over time.

Fig. 113. PXR to Bilateral Peer circuit 2 plus 4 usage over time.

Fig. 114. PXR to Bilateral Peer circuit 2 plus 5 usage over time.
Fig. 115. PXR to Bilateral Peer circuit 3 plus 4 usage over time.

Fig. 116. PXR to Bilateral Peer circuit 3 plus 5 usage over time.

Fig. 117. PXR to Bilateral Peer circuit 4 plus 5 usage over time.