Traffic Congestion Patterns

6/28/1999

Robert Hanmer
Lucent Technologies
hanmer@lucent.com
2000 North Naperville Road
PO Box 3033
Naperville, IL 60566-7033
voice: 630 979-4786

Michael Wu
Lucent Technologies
mwu@lucent.com
2000 North Naperville Road
PO Box 3033
Naperville, IL 60566-7033
voice: 630 979-5615

Abstract

An examination of the successful techniques that have allowed earlier networks and network elements to be successful must be done to understand how to evolve networks. The patterns presented here describe techniques implemented in many Lucent Technologies switching systems to deal with traffic congestion issues. Handling congestion is one of the key roles of a network management system.

Congestion is the term used by telecommunications system designers to describe the situation when the call load exceeds the available resources of the system.

These patterns discuss ways that these systems deal with congestion. The approaches can be generally classified as protective or expansive controls. Protective controls reduce the amount of work by restricting access to resources. This protects the system from too much work. Expansive controls allow the system to use additional resources not available to call processing under normal conditions. This expands the possible actions that the system may take.

Introduction

> “Congestion occurs when the call load exceeds the capacity of available routes or trunks, or of common control equipment (for example, call registers that temporarily store call data while the call is being processed.) Congestion can also occur when the call load cannot be handled in available real time.” [MG]

Telecommunications system designers have studied the problem of system resource congestion for years. These same problems are evident in the distributed and client server computing systems of today. This article describes solutions that have worked in telecommunication systems over time.

Network management functions within telephone switching systems are designed to maintain a high level of service during periods of unusual traffic on systems that are engineered for less than full capacity. During these periods, machine performance naturally degrades due to resource contention and network delays. Even when normal traffic levels are restored the system continues to struggle to achieve normal performance levels. [GHHJ]

Internal congestion arises from two different categories: peripheral capacity and processor capacity. Common control or peripheral equipment capacity is related to how a system is engineered and how requests for service arrive. In the telecommunications world these requests for service are “traffic” and consist of telephone calls being handed off from one system to another as they progress throughout a network. Processor capacity (i.e. processor real time) is taken as the key driver behind a system’s “overload” capabilities. Peripheral capacity issues will be discussed here. For solutions to these problems refer to “A Pattern Language for Improving the Capacity of Reactive Systems” by Gerard Meszaros [Mesz].

We begin by discussing “patterns”, the form in which the solutions we present are described. Then the context of telecommunications traffic and how it appears to switching systems is discussed. This is followed by the actual traffic congestion patterns.

Background

Patterns

This section lays the foundation that is required to understand the patterns presented in the remainder of this paper. A pattern is a description of a good, working solution to a problem that occurs again and again. A pattern contains a statement of the problem, the context where the problem exists, and a consideration of the various trade-offs involved in the solution, as well as the actual solution. Patterns only capture solutions that have stood the test of time and can be used whenever the problem occurs again. Generally patterns must be customized to a system’s unique characteristics to be used.

What is contained in this article are solutions to problems related to peripheral control equipment congestion. These patterns describe functionality needed by a switching system to concentrate upon profitable work by avoiding unnecessary work and to decide in real time how calls are routed in a network. They each solve smaller parts of the larger problem of dealing with congestion, however when taken together they form the nucleus of an effective network management system.

The names of patterns appear as *underlined italics*.

Telephone Switching Systems and Traffic Congestion

A telephone switching system is a system that is intended to connect customer telephone circuits when the customers desire to communicate. These connections are created as a call is being established and are released when the customers hang up. Route selection and the selection of idle bandwidth are routinely done by these systems without human intervention.

Telephone switching systems are real-time, fault tolerant systems that attempt to maintain themselves without requiring constant human supervision (refer to *Minimize Human Intervention* [ACGH+]). But the systems generally consider that sometimes *People do know best* [ACGH+], so provisions for human oversight and altering normal actions are built into the system.

The patterns are primarily drawn from experiences with the 4ESS™ Switch, a high-capacity toll and tandem switch (refer to the *Hierarchical Routing* pattern below for an explanation of these terms). Traditionally the connections between switching offices are referred to as “trunks”. The term “line” refers to the connections between switching offices and individual telephone sets. This pattern language only contains patterns related to congestion on trunks. Generally at the trunk level there are not restrictions that a particular call use a particular route, unlike when a telephone being called resides at the end of only one specific line.

Typically a telephone network requires a large number of switching systems, since individual systems have not been built that have the capacity required to concentrate all of the telephone traffic. Each of these systems is a node in the telephone network. Between these switching systems there are a large number of possible routes.

A large network in which every node is connected to every other node would be infeasible. Every route requires ongoing maintenance of the physical medium as well as maintenance within each switching system node that it connects to. But with more routes in the network, any specific telephone call has fewer switches to traverse; resulting in higher quality and less time required to complete the call.

A balance between number of routes with their associated expenses and the quality of service that is achievable and desirable in the network must be struck. The impact on traffic congestion should be obvious. If every switch is connected to every other with trunks in sufficient quantity then congestion will be non-existent. With fewer links, the possibility of routes concentrating on individual switches increases as well as the probability that that a switch will be subject to control equipment congestion.

Typically these networks have evolved with time. New switching systems replace older, smaller ones. Switching systems are quite expensive, so they are used as long as possible before this replacement occurs. Over time networks become very heterogeneous. This complicates any strategy for dealing with congestion since the switches may not all have the same abilities to control congestion.

The evolution of the network also requires that the systems be designed to be flexible in their routing. The various trunks that are installed in a new switching office might not be the same trunks as when the office is retired or replaced at the end of its useful life.

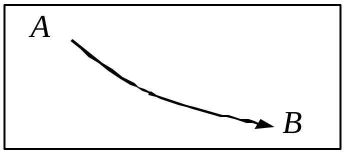
When a telephone network is being engineered usual telephone usage patterns are considered. Since switching equipment is expensive the systems are engineered to take advantage of the fact that each individual telephone line has much idle time. Most customers do not make continuous telephone calls. The systems are designed to allow expensive peripheral equipment to be shared among a number of lines or trunks. This common control equipment can be allocated as needed, and since most lines or trunks are not in use continually this sharing is feasible without diminishing the quality of service. But, shared resources can be needed by too many calls at the same time, resulting in peripheral capacity limitations.

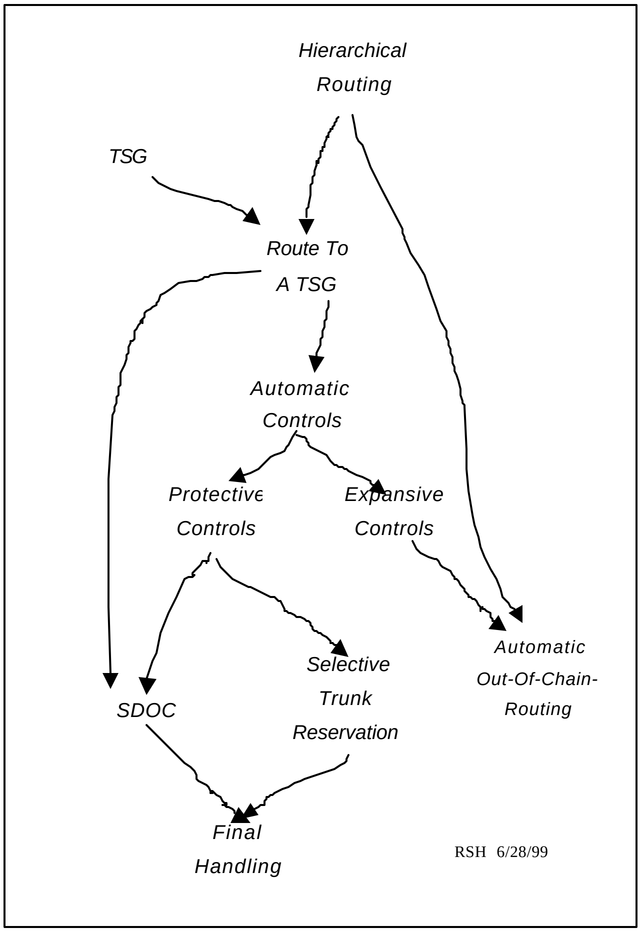
Natural disasters are another complicating factor within the network of switching systems. A common reaction to a disaster is to call your relatives to determine their status. When everyone does this a focused traffic load aimed at switches in the disaster area occurs. This is a kind of peak traffic that the network engineers cannot foresee, and hence frequently causes overload and control equipment congestion.

Many of these concepts: a variety of nodes, infeasibility of directly connecting every switch, and shared resources can be found within the surface transportation network as well as the telephone network. Each of the patterns in this language presents one or more examples from this other network to illustrate the principles discussed by the pattern. These examples also help show the power of these patterns in that they are not limited solely to an electronic communications network.

Language Map

The following diagram shows the relationships between the patterns in this language. The diagram shows patterns that enhance the solutions of other patterns, resolve previously unresolved forces in a pattern, or take advantage of an earlier pattern to provide some new system capability. In the example to the right, pattern *B* refines pattern *A*, helping to solve unresolved forces or new problems that *A* introduced.





The Patterns:

Trunk SubGroup (TSG)

Intent: Group trunks by common properties so that appropriate ones are easy to find.

Problem: From the myriad of different trunks in an office, how do I select the one for this particular call?

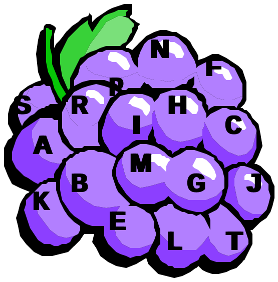
Context: Because of the heterogeneous nature of the switching systems in the telephone network, each is capable of performing call setup signaling with only a subset of the available protocols. Each of these protocols has different advantages and disadvantages for each switch that uses it. Each trunk has attributes related to the protocol used upon it. In addition to this each trunk will have a variety of fixed attributes such as endpoints and targeted traffic.

Forces: A switching network element will have many, many trunks, sometimes up to 100's of thousands. Each of these has several different attributes that can be used to describe them. These attributes are of interest to the routing process so that appropriate classes of service can be allocated to each call traversing the network. Some of the example attributes are:

> Destination, Translation domain, switching domain, traffic use, far end, INWATS, incoming screening class, incoming traffic separation class, directionality, transmission delay, signaling characteristics. [MG]

The system cannot just pick any trunk from the pool of available trunks; it must have the right attributes. This requires that some record be kept of the attributes on a per trunk basis. These records can be used by the routing selection functions to locate the correct one to use.

Solution: When initially engineering the office group trunks into "trunk subgroups" (TSG) based upon common properties. These properties will allow trunk selection (or in the case of Network Management trunk category selection) to be easier since it will be easier to identify similar trunks at route selection time.



Rationale: This helps two aspects of switch application software: Network Management and Call Processing. Network Management is helped because finer divisions allow finer degrees of control. Call processing is allowed to select services by allowing calls to be routed to a particular type of facility.

Resulting Context: Trunk resources with specific attributes can be identified quickly because they are within the same TSG. This simplifies selection and allows the system to ***Route to a TSG*** at execution time.

Highway Example: Associate properties with each highway, such as maximum speed, how direct they are, tolls, etc. These can be used during route planning to provide the desired type of travel experience. Many modern trip assistant software packages actually allow the user to select which categories of roads that they will travel on. This allows the user to find routes with the properties that they desire.

Hierarchical Routing

Intent: Define a fixed way to route calls to neighboring switches

Problem: How do you design your network so that even unsophisticated switching systems can route calls through it automatically.

Context: You can't connect every office to every other. The method has to be easy enough for pre-electronic switching equipment to implement in hardware or with simple wiring changes. You want to design the network to make optimal use of its links and of its common resources.

Forces: If you directly connect every node in a network you have n(n-1)/2 links. From experience we know that there is much expense associated with managing large numbers of links.

The more systems that must process a call, the longer it takes for the call to be established.

Solution: Identify each switch by a class number or level. In the telephone network five levels are used. Class 5 offices are end offices to connect to customer telephone sets. Class 4 offices are called toll offices, class 3's are primary centers, class 2's are sectional offices and class 1's are regional centers. Sometimes local tandem offices connect class 5 offices without the call requiring a class 4 office.

Route each call as low as possible. If the calling and called party are on the same switch, route the call directly. If they are not, go up the hierarchy from the calling party's class 5 office to the lowest switch (highest class number) that can connect the calling party's class 5 office with the called party's class 5. For example the call may be passed through the local office to the toll office to another toll office and back down to the called parties local, class 5 office.

Engineer some links between offices of different classes to provide some shortcuts within the hierarchy.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.