Optimizing Restoration Capacity in the AT&T Network

Ken Ambs
AT&T Network Services, 900 Route 202/206 North
Bedminster, New Jersey 07921

Sebastian Cwilich
AT&T Labs, 200 Laurel Avenue
Middletown, New Jersey 07748

Mei Deng
AT&T Labs

David J. Houck
Bell Laboratories, Lucent Technologies
101 Crawfords Corner Road
Holmdel, New Jersey 07733

David F. Lynch
AT&T Labs

Dicky Yan
AT&T Network Services

To ensure high network reliability, AT&T employs two basic approaches: preventing failures and responding quickly when failures occur. For AT&T to quickly reroute traffic in the event of a network failure, the network must contain sufficient restoration capacity to carry the displaced demand. A team of AT&T OR experts, network planners, and managers developed a method for determining the appropriate quantity and location of restoration capacity required to restore the demand during any single link failure. The approach centers on a linear programming model to minimize the cost of the restoration network and uses column generation to generate new restoration paths as needed. In about 10 months, the team converted the methodology into a tool to optimize the allocation of restoration capacity. This tool was then extended to plan for the recovery of a switching-center disaster and to reoptimize the entire restoration network. It has contributed to AT&T's achieving high-quality service, while saving valuable resources. It resulted in hundreds of millions of dollars in cost savings and increased revenues.

AT&T was incorporated in 1885 in New York as a subsidiary of the American Bell Telephone Company. Its main objective was to manage the long-

0092-2102/00/3001/0026/$05.00
1526–551X electronic ISSN

INDUSTRIES—COMMUNICATIONS
PROGRAMMING—LINEAR—LARGE-SCALE SYSTEMS

distance business of its mother company. AT&T has transformed itself into a global telecommunications company providing a variety of services from long-distance-voice to data, video, wireless, satellite, and internet services. This variety of traffic is carried on one of the largest and most complex communications network in the world, the AT&T Worldwide Intelligent Network. The AT&T network uses state-of-the-art technology to provide switching and transmission services. The network serves over 80 million customers. Within the continental US, the network consists of over 40 thousand miles of fiber-optic cable, forming the largest fiber-optic network in the industry. It handled over 76 billion calls in 1997 with more than 250 million voice, data, and video calls on an average business day. On peak days, AT&T handles as many as 290 million calls. In 1997, 99.98 percent of all calls were completed on the first attempt.

AT&T sees the traffic in the network increasing rapidly and being distributed more and more unpredictably. Because of its scale and coverage, the network's reliability and restoration in case of facility outages are important to business customers, governments, and consumers in their day-to-day use of telephone, credit-card, internet, and commercial services. Various events may interrupt network operations: power outages, equipment failures, natural disasters, cable cuts, and so forth. A failure in the network can create a severe service loss for customers. Network reliability increases significantly when traffic interrupted by failures can be redirected (restored) via spare capacity in other parts of the network.

To ensure network reliability, AT&T employs two basic approaches: preventing failures and responding quickly when failures occur. AT&T has extensive programs in place to prevent failures. However, we focus here on the network's ability to recover quickly when a failure occurs. In particular, we focus on the placement of spare capacity in the network for rerouting circuits around failures.

Maintaining reliability and restoring service in cases of facility outage have become complex business and economics problems. AT&T estimates that its restoration- and reliability-management programs deliver benefits worth billions of dollars to the company annually. We seek to achieve high reliability under tight resource constraints, especially in time, capacity, and budget.

AT&T invests heavily in such systems as RTNR (real time network routing) and FASTAR (fast automatic restoration). In 1992, AT&T introduced RTNR, a patented, highly sophisticated system that detects network conditions and routes voice (telephone) traffic via alternate routes to avoid congested switches and circuits. In addition to RTNR, the AT&T network uses FASTAR to ensure network reliability. FASTAR instantly identifies fiber-optic cable failures on the core network and automatically begins rerouting circuits over spare capacity. Usually, FASTAR restores 90 to 95 percent of the disrupted circuits within two to three minutes, before most customers know there is a problem [Chao, Fuoco, Kropfl 1994].

In 1995, AT&T first introduced SONET (synchronous optical network) technology into the network. It increases network transmission capacity and improves service restoration. A SONET ring is a group of connected SONET facilities deployed in a ring configuration that may be hundreds of miles in circumference. SONET rings employ parallel protection channels to provide subsecond restoration following a fiber-optic-cable failure on the core network. For example, if traffic is moving in a counterclockwise direction on the service channels of the ring when a fiber-optic-cable failure occurs, SONET technology instantly reverses the flow, and the affected traffic then moves in a clockwise direction on the protection channels. This reversal takes approximately 60 milliseconds and does not interrupt calls in progress. Although SONET is generally installed in an interconnecting ring configuration, where a single ring might cover several states, some point-to-point systems are being deployed to provide additional transmission capacity. AT&T has over 50 SONET rings in operation. It achieved coast-to-coast SONET connectivity in 1997.

Another major advance in transport technology is DWDM (dense wavelength division multiplexing), which increases transmission capacity. A typical fiber-optic cable contains 32 or 48 fiber strands, which equals 16 or 24 fiber pairs. In the past, a fiber pair carried up to 24,000 simultaneous calls; however, DWDM allows multiple wavelengths or colors of light to travel a fiber pair, significantly increasing its transmission capacity.

As we enter the 21st century, the telecommunications industry faces increased competition and customer demands, as well as budget reduction and downsizing programs. At the same time, demand for network expansion grows at an unprecedented pace. To be successful, businesses must use all their resources with maximum efficiency. Although AT&T deploys more and more rings in the network, much of the network still uses FASTAR to recover from failures. In the non-SONET-ring portion of the network, the protection capacity can be shared. FASTAR can detect the failures in the network, and by controlling the digital cross-connect systems (DCS) at the nodes, it can reroute the traffic dynamically around the failure. The key to the restoration of the transport layer is that the system must have enough spare capacity at the right locations for FASTAR to find an alternative route in the case of any failure.

Restoration capacity comes with an opportunity cost equal to the revenue that could be generated, far exceeding the construction cost.

Determining the appropriate amount and location of spare capacity is an urgent issue for two reasons. First, as the network grows rapidly to deploy service capacity, restoration capacity must grow at the same pace to maintain quality and reliability. Second, excess dedicated restoration capacity could be used for service, generating revenue. Thus, restoration capacity comes with an opportunity cost equal to the revenue that could be generated, far exceeding the construction cost. Therefore, allocating the restoration capacity optimally has become a challenge for network planners and engineers.

AT&T Network Services (ANS) is the network management and information technology unit of AT&T, responsible for all network operations, including network reliability and restoration. In planning restoration capacity, ANS considers both short- and long-term issues. In long-term planning, it evaluates the impact of making localized changes to the network, such as adding or identifying new restoration routes, removing leased routes, or releasing restoration capacity for service. In short-term planning, it makes annual or semiannual plans for network growth. Typically, a group of planners within ANS prepares a capacity plan for the service network based on a traffic forecast. Based on the service plan, the restoration-network planners devise a restoration-capacity plan. A few iterations between the two groups may be required to reach a plan that satisfies the overall business objectives. The planners then pass the plan on to the organization responsible for implementing it.

In late 1996, ANS asked AT&T Labs for analytical support to improve the overall reliability and cost of the network. From a list of projects considered, they jointly decided to investigate possible improvements in the methodology used to design the restoration network. They chose this project because it was at the core of multiple network issues and would have direct tangible benefits if it made improvements. The goal was either to declare that the existing heuristic for assigning restoration capacity is nearly optimal or to develop better optimization algorithms. A team formed in early 1997 comprised operations managers, network-subject-matter experts, project managers, and researchers from ANS and AT&T Labs. The authors formed the core of this team, and additional people, who are recognized in the acknowledgements, provided assistance.

By the end of 1997, we had developed a new methodology for planning restoration capacity. We built a prototype system and used it for planning several network scenarios. The new system provided significant savings. During 1998, we continued to enhance the model and the system to deal with other important business issues. We completed a more user-friendly system, called RestNet, in the last quarter of 1998.

RestNet has helped AT&T to achieve high-quality service, while saving valuable resources.

Restoration Network Architecture

To deliver results of value, we had to understand network-transmission technologies, the existing restoration systems including FASTAR, and the existing planning tool. The first step we took was to understand the network components, technologies, and terms the planners commonly used (for example, span, RIN, and T3). In AT&T's network, three elements are relevant for restoration planning: (1) the physical layer, (2) the routing element, and (3) the T3 or demand element.

The physical layer consists of fiber spans. A span is a fundamental physical component in the network connecting two terminals. A span may fail because of various causes, including cable cuts and natural disasters. The key information about a span includes its identification, termination points, and length.

The routing element consists of a series of spans of a common transport technology (typically fiber-optic cable) with uniquely defined end points, which are usually nodes containing a digital cross-connect system (DCS). In the AT&T network, routing elements are called route-identification numbers (RINs). (In spite of its name, a RIN is actually identified by an alphanumeric code.) FASTAR can access restoration capacity on RINs and automatically reroute traffic around failures. Routing cannot be changed at intermediate points of a RIN however; a demand must traverse an entire RIN.

A T3 (45Mbps) is a basic unit of demand for which the transport network is engineered and designed. Each T3 is routed over a sequence of RINs. The T3s are either individual demands of 45 Mbps or they are composed of a set of smaller demands, such as T1s (1.5Mbps) riding on that T3. FASTAR restores at the T3 level; it does not restore individual smaller demands.

AT&T uses a restoration tool that models FASTAR and the dynamic routing algorithm RTNR, and that, for any span failure, can provide the corresponding set of failed T3 demands and estimate the number of blocked calls in the voice network. Prior to our work, AT&T also used this tool to allocate the restoration capacity in the network. The heuristic employed by this tool considers failed spans one at a time. For each demand affected by the span failure, the algorithm finds an alternate path that requires minimum spare capacity and then allocates enough spare capacity on the alternate path. The same spare capacity may be used by two or more span failures. The tool repeats this procedure for all spans. The order in which the algorithm considers possible span failures can influence the allocation of spare capacity in the network.

Problem Description

Assigning spare capacity is a special network-design problem in which spare capacity must be installed in a network to insure quick restoration of disrupted services in the event of any failure. Restoration capacity accounts for a large part of the infrastructure cost of telecommunications networks, so allocating restoration capacity is a major task for network planners. The objective of the network-restoration problem (also known as the spare-capacity-assignment problem) is to determine where and how much spare capacity to install in a network while minimizing facility cost. Our task was to find the minimum restoration capacity necessary to restore the given set of demands for all possible single span failures.

To understand the network-restoration problem, consider a small abstraction of a transport network containing a route between Cincinnati, Ohio and Middletown, New Jersey (Figure 1).

Suppose that a backhoe digs up the link between Parkersburg, West Virginia and Pittsburgh, Pennsylvania (Figure 2). FASTAR-capable nodes automatically detect and locate the failure and send the information to FASTAR. FASTAR identifies the T3 demands that are affected, finds restoration routes for each one sequentially in priority order, and sends orders to the nodes to make the appropriate connection changes, establishing a restoration

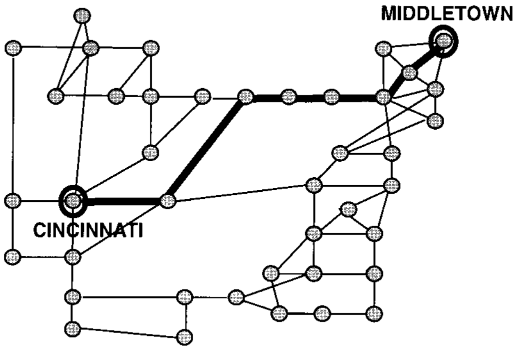

Figure 1: This portion of a telecommunications network contains a connection between Cincinnati, Ohio and Middletown, New Jersey.

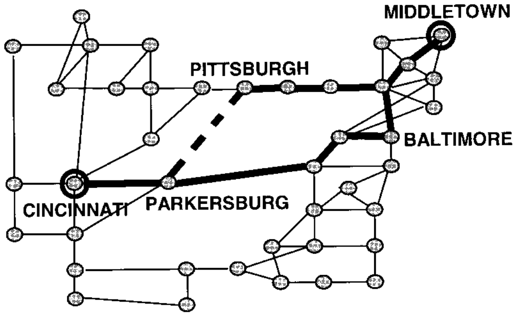

Figure 2: If a span between Parkersburg, West Virginia and Pittsburgh, Pennsylvania fails, all traffic using that span must be rerouted over other portions of the network. In this example, traffic from Cincinnati to Middletown is rerouted through Baltimore.

route for the Cincinnati to Middletown traffic through Monrovia, Finksburg, and Baltimore, Maryland.

This is just one of possibly hundreds of demands between different pairs of nodes in the network that use the Parkersburg to Pittsburgh link. We want to be able to reroute as many demands as possible (using different routes for different demands) and place capacity in the network to plan for every possible span failure that might occur. Our task is to determine how much restoration capacity to put in the network and where to put it to achieve the same high level of network reliability currently in the AT&T network. The same restoration capacity can be shared among several span failures. The network that we are dealing with is quite large with hundreds of nodes, thousands of RINs, and tens of thousands of demands.

Linear Programming Model

Several researchers have previously examined the network-restoration problem using different models and assumptions. (Cwilich et al. [1999] did a detailed survey, and a new model was recently proposed by Kennington and Whitler [1999].) Because the problem is a global-network-optimization problem and has a very nice structure, it is natural to formulate it using linear programming (LP). Two types of network models use linear programming: arc-path-based formulation and node-arc-based formulation. We chose the arc-path formulation because it results in a smaller number of constraints for this model.

Each restoration-capacity plan specifies where and how much to augment the existing restoration network. The LP model must accommodate two types of RIN augments: restoration-capacity augments and SONET spare-capacity augments. The SONET spare capacity in point-to-point SONET links or SONET linear chains is available for protection of that link or chain. However, to be used by FASTAR in automated restoration, it must be cabled up to a DCS at each end-point node. We refer to this capacity as preinstalled RIN capacity. The cost of cabling up a preinstalled RIN is less than the cost of building a new restoration RIN, and that cost is reflected in our model. The number of preinstalled RINs that have not yet been cabled to a DCS bounds the capacity for preinstalled RINs. In addition, the idea of preinstalled RINs can be extended to optimize the embedded capacity base in the network. In that case, instead of the embedded base having zero cost, it is included as pre-installed RINs with an appropriate cost (relative to new builds) and upper bounds denoting the available capacity.

The linear programming model can be summarized as follows. Minimize the cost of RIN augments and the preinstalled RINs to be cabled, given that we must restore the failed demands. We assign a small cost for a preinstalled RIN and a large cost for a RIN augment. The cost of a preinstalled RIN should be proportional to the cost of cabling up a preinstalled RIN and the cost of an augment is proportional to the cost of building a new RIN in the network. We use the ratio of 0.05 between the two costs for most of our studies.

There are two sets of constraints:

—Demand constraints: to restore as many failed demands as possible for each possible span failure; and

—Capacity constraints: for any span failure, the total usage by all restoration paths on a RIN cannot exceed the total restoration capacity available on this RIN, which is the sum of the embedded capacity, the preinstalled capacity, and the new augments.

We give the mathematical formulation in the appendix.

The solution of the LP model provides the minimal cost combination of preinstalled RINs and new RINs required to restore as many failed demands as possible. Since it is not always possible to restore all demands because of network topology and individual RIN limits, we included dummy paths to accommodate this situation.

Although the arc-path formulation results in a smaller number of constraints than the node-arc-based formulation, the network that we are dealing with is quite large, with hundreds of nodes, thousands of RINs, and tens of thousands of demands. The size of the network can cause the LP model to have millions of variables and constraints.

Modeling Methodology

Our modeling methodology consists of seven steps: defining the problem; aggregating spans, RINs, and demands; generating paths; solving the LP; integerizing the solution; disaggregating the problem; and delivering the plan to the organization responsible for implementing it (Figure 3).

Problem Definition

To define the problem, we employ the existing restoration tool to find the set of demands (the T3s) that need to be restored for each possible span failure. We also use this tool to provide a consistent description of the network topology (nodes, spans, and RINs), the routing of existing demands, and the embedded restoration capacity. Other planning inputs include a list of capacity-constrained nodes at which no new augments could be built during the current planning cycle, potential cost models of different augments, and links with limited amounts of capacity augmentation available.

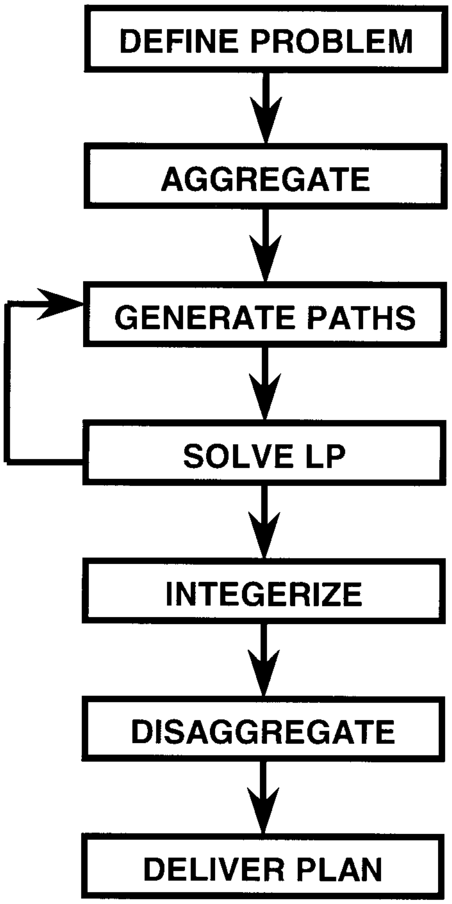

Figure 3: We use a seven-step process to design the restoration network: defining the problem; aggregating spans, RINs, and demands; generating paths; solving the LP; integerizing the solution; disaggregating the problem; and delivering the plan to the organization responsible for implementing it.

Data Aggregation

To reduce the size of the problem to be solved and thus to speed the solution process, we aggregate the network data using three steps: RIN aggregation, demand aggregation, and span aggregation.

We exclude any RIN terminating on a node that does not contain a DCS, since these nodes cannot perform automatic restoration, and then aggregate parallel RINs that ride on the same physical route into a single RIN.

We map the original RIN path of each demand to the aggregated RINs and condense intermediate RINs to those with a DCS in each node, introducing new pseudo RINs where necessary. We then combine any demands that are routed on the same set of aggregated RINs into a single aggregated demand.

We combine spans that carry the same set of aggregated demands and (when failed) affect the same set of RINs.

Initial Path Generation

Although the LP model can provide an optimal solution for RIN augments, FASTAR uses a heuristic algorithm to compute restoration paths in real time. To improve the convergence of the LP and to consider restoration paths favorable to FASTAR, we start with an initial set of paths chosen by a heuristic that approximates what FASTAR would do. For each span failure and for each T3 affected by that span failure, we find the shortest path using embedded capacity. If there is no such path, we find a path that needs the smallest number of RIN augments. We add the new RIN augments to the set of embedded capacity, making it available for the next span failure.

Using Column Generation to Solve the LP Model

For most of the restoration problems, even after aggregation, there are still hundreds of RINs and nodes, and tens of thousands of demands. It is computationally impossible to generate all feasible restoration paths for every affected demand. Even if we could generate all of the feasible paths, the size of the resulting LP model would be too large to handle.

To find the optimal solution for the LP model, we implemented a block-mode column-generation technique. The column-generation method repeatedly solves a linear program to choose restoration paths for each demand and each possible span failure from a set of candidate paths. The key idea is to generate paths only when needed. The total number of restoration paths generated could be in the millions.

In column generation, after solving an LP, we use the dual variables from the current solution to generate better candidate restoration paths for the next iteration. For each combination of span failure and demand, there will be a set of dual variables corresponding to each RIN in the network. We find the shortest path using Dijkstra's algorithm [Lawler 1976], connecting the origin and destination of this demand, using the dual-variable values as costs. This provides a restoration path for that span-failure-and-demand combination that has the minimum reduced cost among all possible paths. By doing this for each span-failure-demand combination, we obtain a set of restoration paths that we might use. Adding these paths to the previous set and resolving the LP may improve the solution. If the cost of the shortest path for any demand is equal to the dual variable corresponding to the demand, then there are no better paths for that demand (using the current basis). When no better paths can be found for all demands, the column generation process ends and the problem terminates at an optimal solution.

We give the dual formulation of the LP model and discuss column generation in the appendix.

Dynamic Path Control Policy

Although the process of column generation in general converges to an optimal solution, it may take a very long time. The convergence speed is crucial to the performance of column generation.

Based on our experience, if we keep a small number of paths per demand per iteration, the convergence speed of column generation may be slow because some paths may be regenerated at a later iteration. On the other hand, if we retain all the paths generated, for a large problem (where millions of paths may be generated), it may take a long time to solve the LP problem for a single iteration. Hence, the total convergence speed could also be slow. To overcome the convergence problem, we introduced a dynamic path-control policy. We borrowed the idea from the famous s-S policy for inventory control.

We set an upper bound S, the maximal number of paths retained in the problem, and a lower bound s, the minimal number of paths retained in the problem. If the total number of restoration paths generated is less than or equal to the upper bound S, then we keep all the paths. If the total number of restoration paths generated is greater than S, then we delete some paths so that the total number of paths is equal to s.

The criterion for deleting paths is the reduced cost associated with each path from the LP. Generally, the reduced costs are ordered from small to large. We retain the top s paths (from the smallest reduced cost) in the problem.

How to set S and s is a separate optimization problem. In our practice, we use some heuristics to set the parameters.

Also, we can generalize this idea to control the number of paths kept per demand by merely setting S_t and s_t for each demand t.

Integerization

For many of the restoration-planning problems we solved, the LP solutions turned out to be mostly or entirely integer in terms of RIN augments. We could easily handle the few instances of fractional RIN sizes through rounding. For the basic model described so far, we use the LP solutions only for network sizing and not for restoration routing since FASTAR finds its own restoration routes in real time. Thus, it does not matter whether the path variables are integer as long as the restoration system (FASTAR in the case of the AT&T network) is able to find feasible restoration paths. Nevertheless, it is still possible that an optimal solution from the LP model will end up with many fractional RIN augments. In fact, as we expanded our set of test problems, we found cases in which over half of the RIN variables in the LP solution were noninteger. Consequently, a method for obtaining integer capacity values is necessary.

The key idea of the capacity-integerization algorithm is to gradually round the fractional values to integers while minimizing the total cost. We have found in practice that this technique produces a feasible integer solution at a cost that is usually within one percent of the cost of the (optimal) LP solution.

Sometimes we would like to obtain a set of restoration paths directly from the LP solution. This may be because we want to implement predetermined restoration paths in the network, or it may be because we simply want to verify the LP output. Rounding up the path variables to the next integer will not always result in a feasible solution. Instead, we use a heuristic approach to obtain feasible restoration paths from the LP output.

This path-integerization routine always starts from an integer capacity solution. It is composed of two major steps. In the first step, we apply a rounding technique to all of the paths with a fractional value. Rounding a path variable up could result in rounding another path variable down. Second, we implement a path-generation routine for all the demands that still need some restoration paths after the first step. The restoration paths with the minimal number of additional augments from the integer capacity solution are preferred. Experience has shown that this algorithm works very efficiently. It has found feasible integer path solutions with very few additional augments from the integer capacity solution for all of the problems we have solved so far.

Disaggregation

Before the solution could be used, we had to reverse the aggregation steps. Our initial aggregation methods were very aggressive in reducing the size of the linear program, and it was possible to encounter some situations that could not be disaggregated exactly. Revising the aggregation process made disaggregation straightforward.

Implementation and Results

After we obtained a solution from the LP model, we faced two big concerns, and the answers to them would determine the fate of the approach. The first concern was whether the LP solution would be better than the solution provided by the existing planning tool. The second concern was whether FASTAR could use the RIN augments recommended by the LP. Since FASTAR has its own routing algorithm, it was not clear that FASTAR would be able find routes in the network designed by the LP.

Developing the model and the methodology took most of 1997 from March on. Because AT&T was making major investments in the network, it wanted to see results or at least know whether this approach was going to work. Around June, AT&T was to begin planning for the 1998 network and urgently wanted some answers. However, our prototype software was not yet ready, much less a formal planning tool. We tested the validity of our approach in several ways. We solved small subnetwork problems before the aggregation module was ready. We validated the results with a FASTAR simulator to prove their potential benefits. We also set improvement expectations at five to 10 percent and ascertained that improvements in the solution smaller than five percent would be worthwhile.

The AT&T restoration network is designed incrementally, which means that periodically AT&T adds capacity to the existing base of restoration capacity. The network is composed of hundreds of nodes and thousands of links. Aggregation has the effect of reducing the number of links (both by eliminating parallel arcs and by removing non-FASTAR-capable links) so that we are left with hundreds of links instead of thousands.

We used the already existing 1997 plan as the first major test and in August finally solved the complete problem and demonstrated significant improvement (Table 1). We compared the RestNet method to the heuristic method used in the existing restoration tool. For the 1997 incremental plan, the LP solution required 39 percent fewer RIN augments! In addition, we found a complete integer augment and integer path solution without any increase in the objective value. The disaggregation step for this problem turned out to be trivial. We also solved the unaggregated version of the problem to test our code and the aggregation procedure. It generated the same number of augments as the aggregated version. This confirmed the validity and the value of our methodology.

We then moved on to the 1998 network plan that was already being developed by the regular team using the existing tool. For the initial 1998 incremental plan, the

				RIN Augments	
	Demands	LP Rows	Paths Generated	RestNet	Previous Method
1997 Plan	20,000	30,000	55,000	675	1,114
1998 Plan	25,000	40,000	70,000	1,465	2,300

Table 1. This table summarizes some key characteristics of two typical network-restoration problems solved with our methodology. The restoration-network designs produced by RestNet require 36 to 39 percent fewer RIN augments than the designs produced by the previous method. All quantities represent aggregated quantities.

	Revised Incremental Plan	Full Network Plan
Embedded capacity	9,182	2,353
SONET spare capacity	8,682	9,923
New builds	874	874

Table 2. AT&T used RestNet to reoptimize the entire 1998 restoration network. The second column corresponds to the revised 1998 network incremental plan (new demand forecast and revised network description). The last column shows that with no additional new builds and a small increase in cabling of SONET spare capacity, we returned over 6,800 embedded restoration T3 segments to revenue-generating service! All quantities represent number of RIN augments.

RestNet solution showed a saving of over 36 percent. This was consistent with the saving for 1997. We validated the RestNet solution using a FASTAR simulator and showed that it met the same performance objectives. Thus, the plans RestNet designed are usable by FASTAR. This demonstrates that we can maintain the same network reliability by installing a smaller amount of restoration capacity.

When we ran the much smaller 1997 problem for testing purposes in unaggregated form, it took six CPU days, compared with two hours for the aggregated version. The platform we used for these runs consists of a SUN Ultrasparc machine (a SUN enterprise 3000 server) with approximately two gigabytes of memory, and we used CPLEX version 4 as the LP solver.

The number of columns in the LP depends on the number of paths generated between column-generation iterations, but it is interesting how efficient our approach can be. For runs similar to those for the 1997 and 1998 plans, despite the vast number of possible paths that could be generated for each demand, we found that the average demand required generation of fewer than 10 paths before the LP converged (fewer than three paths for the 1997 and 1998 plans).

Additional Applications

Because of its success, our approach generated lots of interest and ideas for changes, new features to add to the model, and new problems to solve. Among them, the full-network-optimization problem topped the list. The full-network problem called for optimization of the entire embedded restoration capacity. The objective is to release as much embedded restoration capacity as possible while maintaining the same network reliability. In solving it, we do not just search for the minimal extra capacity necessary to restore the demands, but rather we input the entire embedded base as preinstalled RINs for the LP to optimize along with the regular new RINs. This problem posed many technical challenges; it was an order of magnitude larger than the previous problem.

This problem is of particular business interest because any embedded capacity identified as unnecessary by the RestNet algorithm can be returned to the service network to bring in additional revenue (Table 2).

The 1998 restoration plan was the first application of our methodology. With no additional new builds and a small increase in cabling of SONET spare capacity, we returned over 6,800 embedded restoration T3 segments to revenue-generating service!

Disaster Recovery Model

Calls traveling the AT&T network are automatically routed to their destinations by 4ESS switching systems developed by Lucent Technologies. The domestic network uses 136 high-capacity 4ESS switches to automatically route calls to their destinations.

In addition to short-term facility failures, AT&T must protect itself against long-term disruptions caused by such disasters as fires, hurricanes, or earthquakes. The AT&T network contains an extra 4ESS switch, called the disaster-recovery switch. Normally, this switch is used only as a tandem switch to interconnect other 4ESSs. However, in the event of a disaster at one of the other 4ESS locations, AT&T would truck portable DCSs into the disaster location and rehome the access trunks from the disaster site to the disaster-recovery switch (Figure 4). Our task was to determine the amount of restoration capacity the network would need to recover from any single 4ESS failure.

To further complicate the situation, there are restrictions on which restoration facilities can be used for disaster recovery. Restoration capacity can be used for disaster recovery. However, SONET spare capacity cannot be used on a long-term basis because it is needed for periodic maintenance and span protection.

The basic model can easily be extended to plan the capacity necessary to recover from a switching-system disaster. We ap-

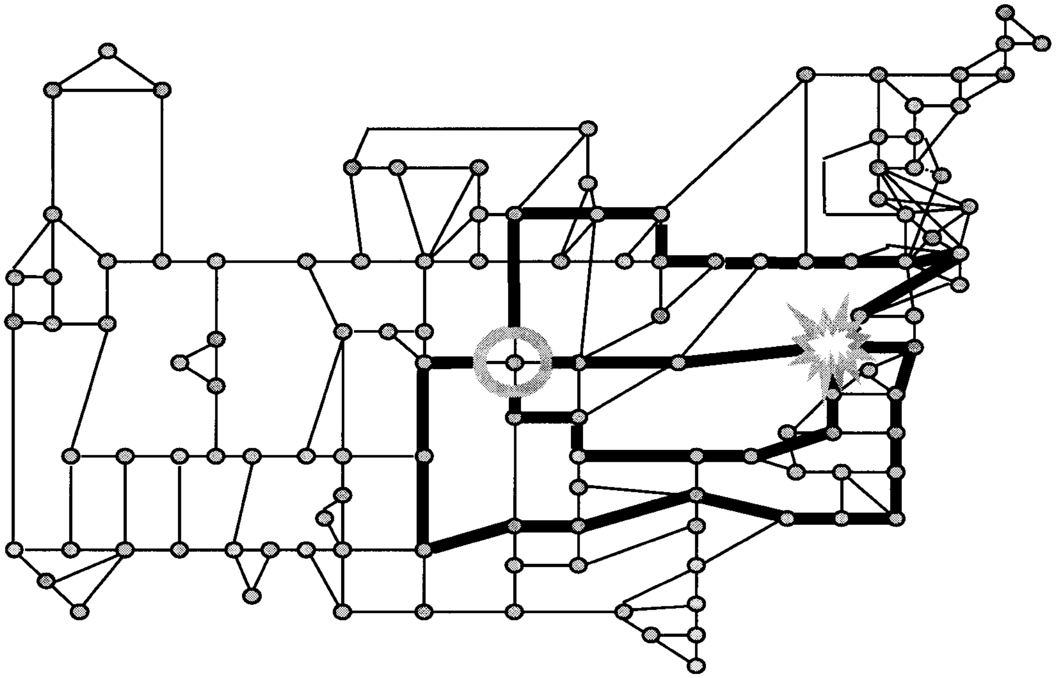

Figure 4: If a switching system in the network fails because of a disaster, all of the customers connected to that switch must be rerouted to a disaster-recovery switch, indicated by the circle.

proached this problem by adding a set of disaster-recovery spans, one for each possible switch failure. We defined a set of disaster-recovery RINs on top of the disaster-recovery spans, mapping disaster-recovery RINs one-to-one to disaster-recovery spans. Each disaster-recovery span carries one demand from the affected switch site to the disaster-recovery switch site, with the amount of capacity that must be rehomed from the failed switch to the disaster-recovery switch.

The LP model for disaster recovery is similar to the basic model except that it includes three sets of preinstalled capacity variables and three types of embedded capacity used by different spans. There are several versions of this problem, and the mathematical formulation of one of them is in the appendix. To solve the disaster-recovery model, we use the same column-generation approach used in the basic model.

For the disaster-recovery model, we use a maximum-flow algorithm to obtain the restoration paths for the traffic from each switch to the disaster-recovery switch. For each span failure, there is a single source-and-destination pair (the failed switch and the disaster-recovery switch). Thus, we can find the restoration paths by solving the single-commodity maximum-flow problem using the given restoration capacity in the network.

Before releasing the 6,800 embedded restoration T3 segments to service (Table 2), we wanted to make sure that enough embedded and spare-service capacity would still be available for disaster recovery. We built the disaster-recovery scenario on top of the revised 1998 incremental plan but without the 9,182 embedded restoration RINs. Therefore, we reduced the number of T3 segments that could be released to 5,000.

Linear Chain Model

AT&T deploys much of its current AT&T network on SONET technology. In the transport network, an optical-fiber route often connects a chain of nodes. To avoid the cost of unnecessary terminations, AT&T divides the fiber channel into a single express RIN that runs the entire length of the fiber route and a chain of local RINs that connect the nodes on the route. In some instances, there may be multiple express RINs of different lengths (Figure 5). To determine the best mix of express and local RINs along each of the linear chains, we modified the basic model to reflect the situation.

We added a new set of linear-chain con-

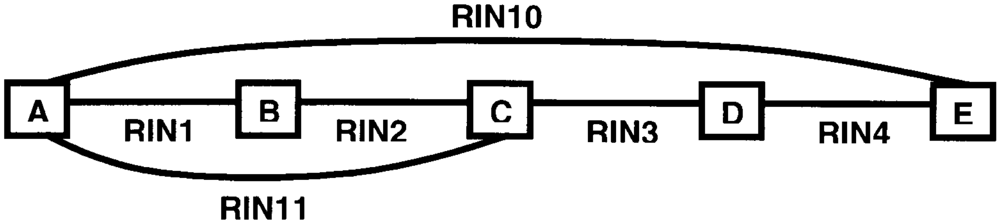

Figure 5: A fiber-optic cable can contain a combination of local and express RINs. In this example, RINs 1 through 4 are local, while RIN 10 runs express from A to E and RIN 11 runs express from A to C. The basic model can be extended to find the best mix of local and express RINs.

straints to the model. Basically, we put a set of capacity-conservation constraints along the local RINs. For example, in Figure 5, we added four linear-chain constraints to the LP model. In the first linear-chain constraint, we bounded the sum of the capacity used by RIN1, RIN10, and RIN11 by the maximum capacity available between node A and node B. In the last linear-chain constraint, we bounded the sum of the capacity used by RIN4 and RIN10 by the maximum capacity available between node D and node E.

To solve the linear-chain model, we used a column-generating approach similar to that in the basic model and the same integerization method.

The linear-chain model can be used to alleviate the DCS port-exhaustion problem. A limited number of ports are available on a DCS, which limits the number of RINs that can be dropped. By running the linear-chain model, we can identify the minimum number of RINs that must connect to a DCS and release the remaining ports (reducing costs in the meantime) for future use.

Link-Disjoint Restoration Model

AT&T is moving forward to a next-generation network that relies on new optical-layer restoration techniques. As part of a study to determine the future direction of the AT&T network architecture, we used RestNet to design the restoration network for various architecture proposals.

To evaluate some of the proposals, we used the basic restoration model using OC48s instead of T3s. (An OC48 is a 2.5 Gbps signal capable of carrying 48 T3s.) However, some of the proposals we evaluated assumed optical-switching technology in the network in which faults could be detected only at the end points of a demand. Because the specific link causing the failure would not be known, the spans in the restoration path would have to be completely disjoint from all of the spans in the service path. Thus, we would need to determine only one set of restoration paths for each demand, but each restoration path would have to be feasible for all possible span failures affecting the demand. We refer to this scenario as the link-disjoint restoration model.

In the LP formulation, the path variables are no longer span dependent. The resulting model has a smaller number of demand constraints, but they are no longer separable by span failure. The rest of the model is similar to the basic model.

Because the LP formulation for this particular problem is not separable by span failure, the column-generation approach is somewhat different from that used for the basic model. Although we can still separate dual variables for RINs by span as in the basic model, we obtain the cost for a RIN in path generation by summing the dual variables that correspond to the spans that affect a demand. In addition, we must modify the path-integerization routine.

In computational tests, the link-disjoint model required only about two percent more restoration capacity than the basic model. Thus, although it imposes significant additional constraints on the basic model, the link-disjoint model requires relatively little additional capacity.

This model can be easily extended to generate node-disjoint restoration paths.

Conclusion

The network-restoration problem is an important but difficult problem. The linear-programming-based methodology has proved to be efficient. The reduction in problem size due to aggregation, the ability to use linear programming instead of integer programming, along with a column-generation approach that has proved efficient at generating a relatively small number of good paths from the vast possible set of all paths allows this approach to be used in large real-life networks. Although we focus mainly on restoration in cases of link failure, we could also develop an LP-based model for node failures.

RestNet has decreased the network resources required to maintain existing levels of service. Compared with the previous methodology, it saves over 35 percent of the spare capacity, capacity AT&T needs to meet growth in the network. For the 1998 growth plan, this translates into savings of 800 RIN augments, or tens of millions of dollars of capital that would have been spent if the augments were constructed from scratch; this excludes savings in maintenance and other costs. AT&T can use the saved resources to build service capacity that generates an average annual revenue of tens of millions of dollars.

We also used RestNet to reallocate the embedded restoration capacity in the network. After meeting increased reliability needs for disaster recovery, the RestNet plan calls for a release of over 5,000 T3 segments of restoration capacity. These represent an average annual revenue impact of hundreds of millions of dollars, even if only half of them are put to service.

Easily overlooked are the time savings. The time that would have been required for building bigger and more expensive restoration plans can now be used for addressing other urgent operations issues. Furthermore, if the growth plan had called for capacity that could not be deployed in time to meet customer demands, AT&T would have suffered lost revenue, contractual penalties, and bad publicity.

We have introduced three extensions to the basic network-restoration model and discussed methods for determining integer solutions based on the LP output for each of the models. The extensions address important issues, such as disaster recovery, DCS port exhaustion, and next-generation network architecture. The link-disjoint model can be easily extended to a node-link-disjoint model, where the nodes as well as links of the restoration path must be completely disjoint from the service path. Furthermore, it is fairly straightforward to combine two or more of these models to address a specific network situation.

RestNet has made major contributions to AT&T's business. AT&T has adopted it as its official restoration-planning tool and has filed a patent application for this innovative application of operations research to business. From a practical economic viewpoint, the approach has produced economic benefits worth hundreds of millions of dollars through the strategic allocation of critical network resources.

Acknowledgments

We thank Clayton Lockhart, Percy Tarapore, and Luis Vitoria for their continuous support and helpful discussions, comments, and suggestions. We express our sincere thanks to Jim Schmitt for his pioneering work on this project. Quynh Nguyen assisted us with the maximum-flow code used to integerize the disaster-recovery model. Steven Phillips and Jeff Westbrook reviewed our work and identified some improvements. Finally, Tom Afferton, Bruce Cortez, and John Strand identified the need for a link-disjoint model.

Executive summaries of Edelman award papers are presented here. The complete article was published in the INFORMS journal Interfaces [2000, 30:1, 26-44]. Full text is available by subscription at http://www.extenza-eps.com/extenza/contentviewing/viewJournal.do?journalId=5