What can be saved with limited recovery resources in a largely damaged optical network exploiting service-level differentiation?

Sedef Savas

Feb. 26, 2016

Networks Lab Group Meeting

University of California, Davis

Introduction

Large-scale post-disaster scenario:

- Multiple network elements fail simultaneously or sequentially.
- Many connections got disrupted, virtual networks got disconnected.
- Limited: **resources left**, **available recovery teams**, **physical tools/hardware**

In a post-disaster scenario, the main objective should be to get the most out of the **remaining resources** through software-level recovery (reprovisioning) while increasing resources through physical-level recovery (repairs) as quick as possible.

In this work, we propose **Dynamic Network Recovery** (DNR) with limited recovery resources to provide an acceptable level of service as much as connections as soon as possible.

Related work

Similar classic optimization problems:

- prize collecting traveling salesman problem (Balas, 1989),
- the vehicle routing problem (e.g., Laporte, 1992; Toth and Vigo, 2002),
- the prize collecting Steiner tree problem (Goemans & Williamson, 1995),
- the orienteering problem (Golden, Levy, & Vohra, 1987),
- ambulance routing,
- road network recovery (for humanitarian response),
- generally the problems of the fields of operations research, managements sciences.

Related work (cont.)

- On progressive network recovery after a major disruption (Chunming Qiao -2011 infocom) – set covering problem. No connections.
- Progressive Recovery For Network Virtualization After Large-Scale Disasters Nasir Ghani (ICNC 2016) – (Qatar National Research Fund) – runs VNE at each stage

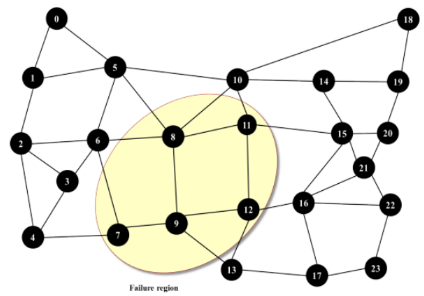


Fig. 7. Cloud datacenter and network substrate topology

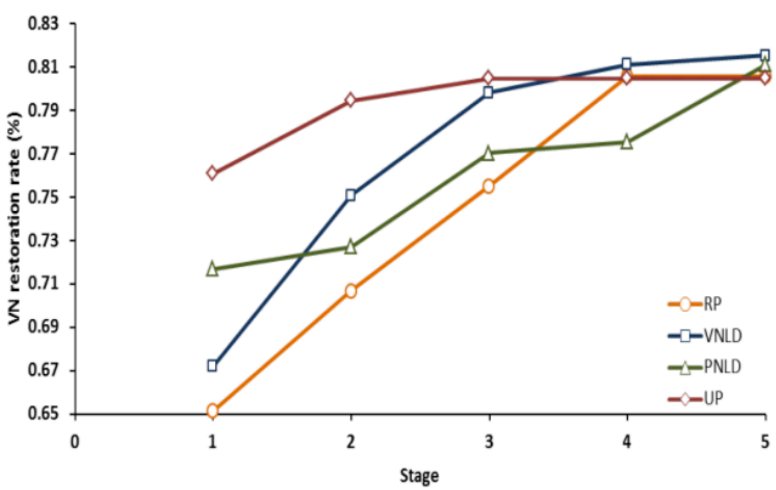


Fig. 8. VN restoration rate

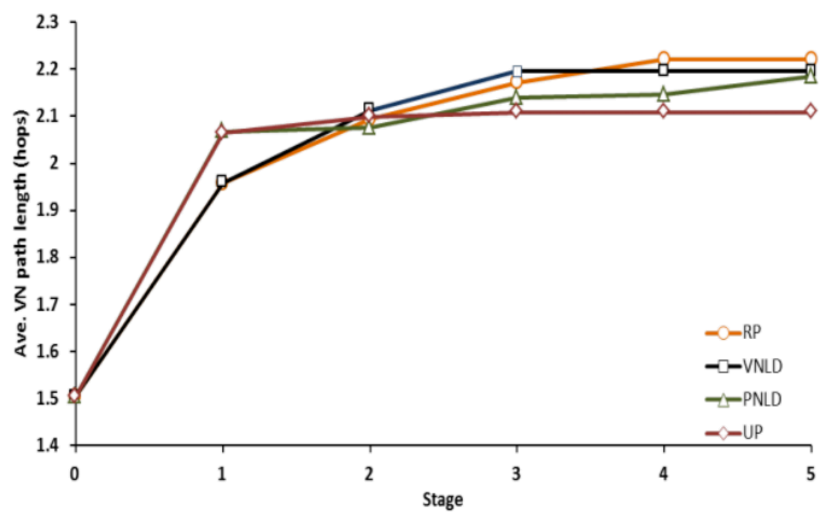


Fig. 9. Average VN path length

Related work (cont.)

Multiple Traveling Repairmen Problem with Virtual Networks for Post-Disaster Resilience (Chen Ma – ICC 2016) – also considers physical links – different vehicles – multiple failures per link – reprovisioning.

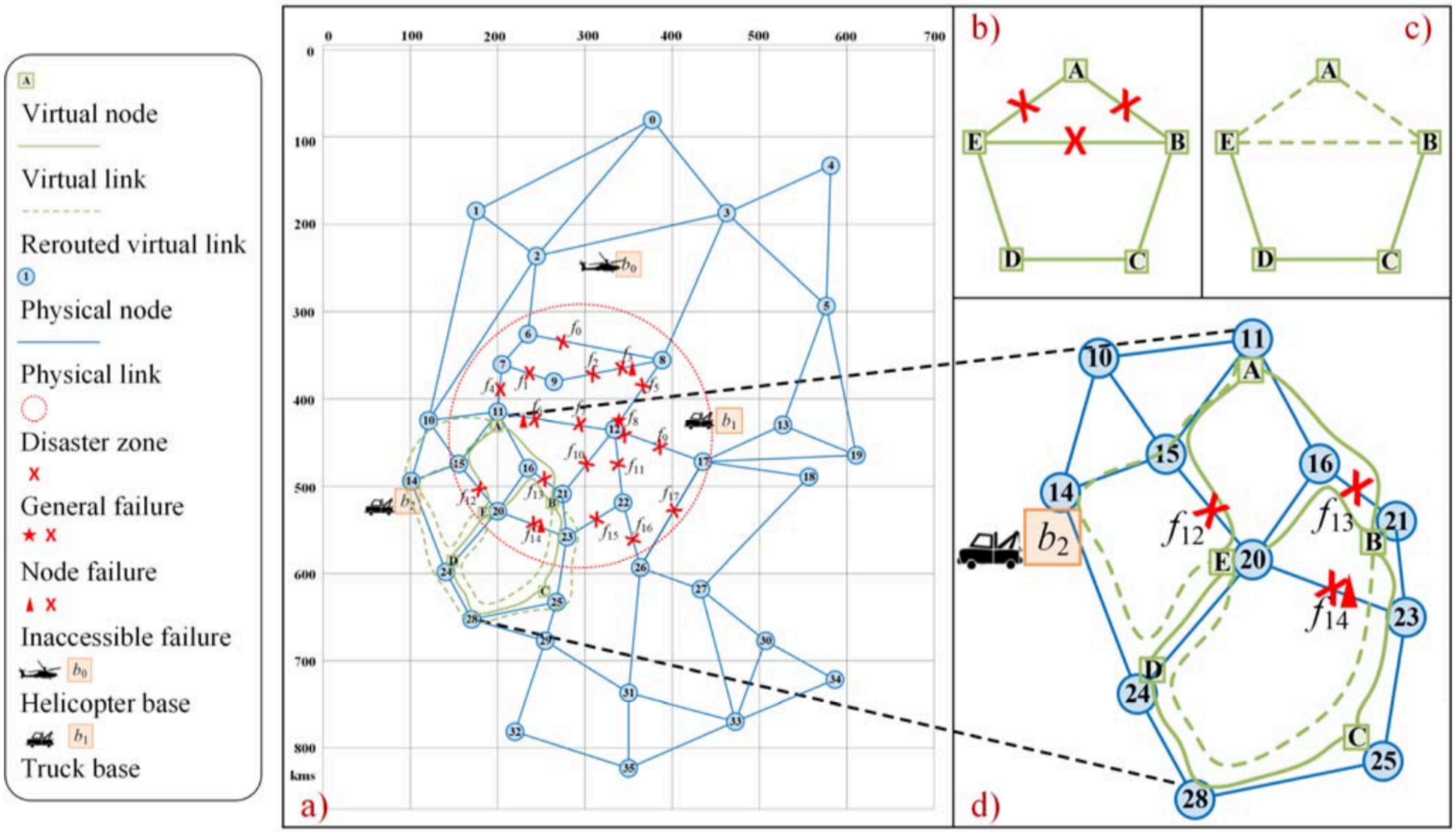


Fig. 1 Illustration of MTRP with re-provisioning. (a) Illustration of MTRP in a German-wide telecom network; (b) virtual network G_0^V when disater occurs; (c) virtual network G_0^V after virutal link re-provisioning; and (d) details of mapping and re-provisioning for G_0^V.

Related work (cont.)

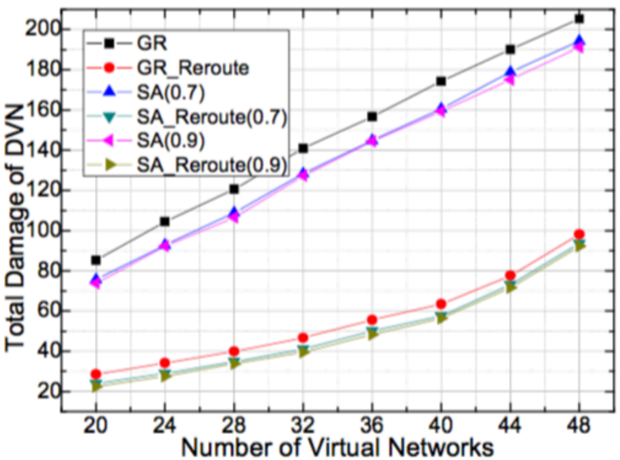


Fig. 4 Total damage of DVN.

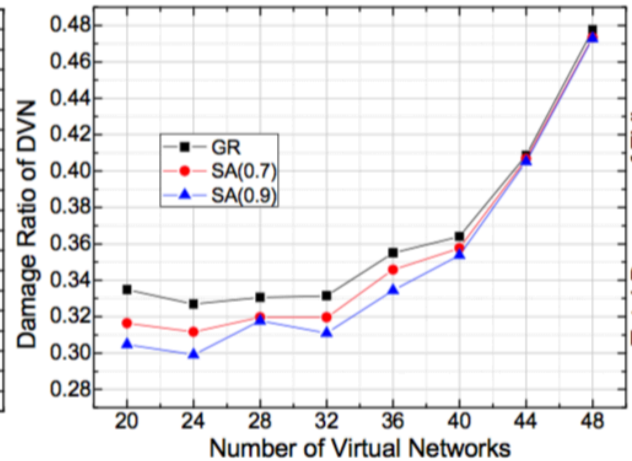


Fig. 5 Damage ratio of DVN.

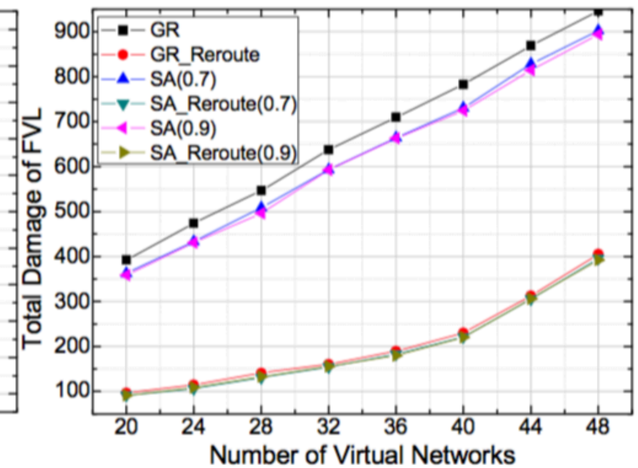


Fig. 6 Total damage of FVL.

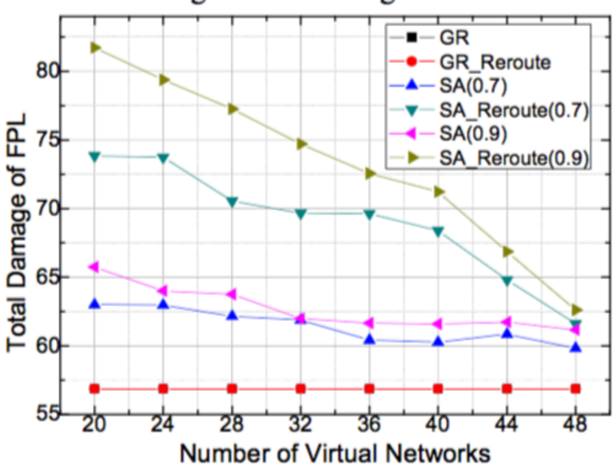


Fig. 7 Total damage of FPL.

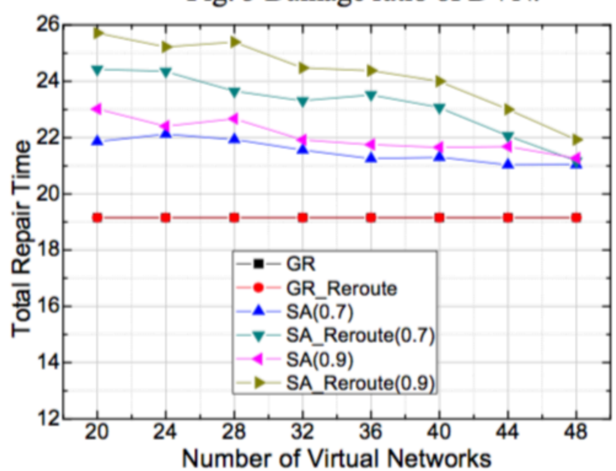


Fig. 8 Total repair time.

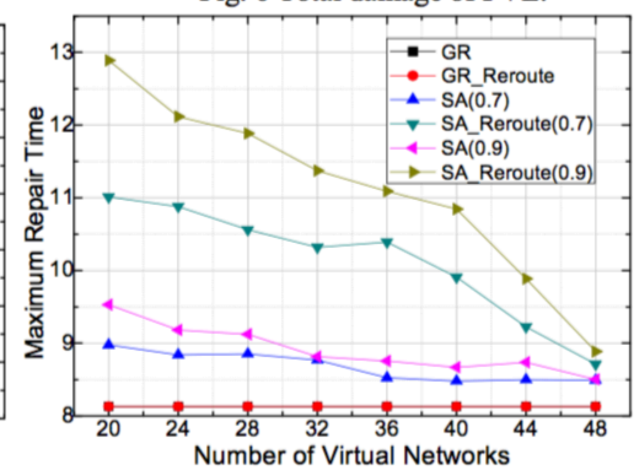
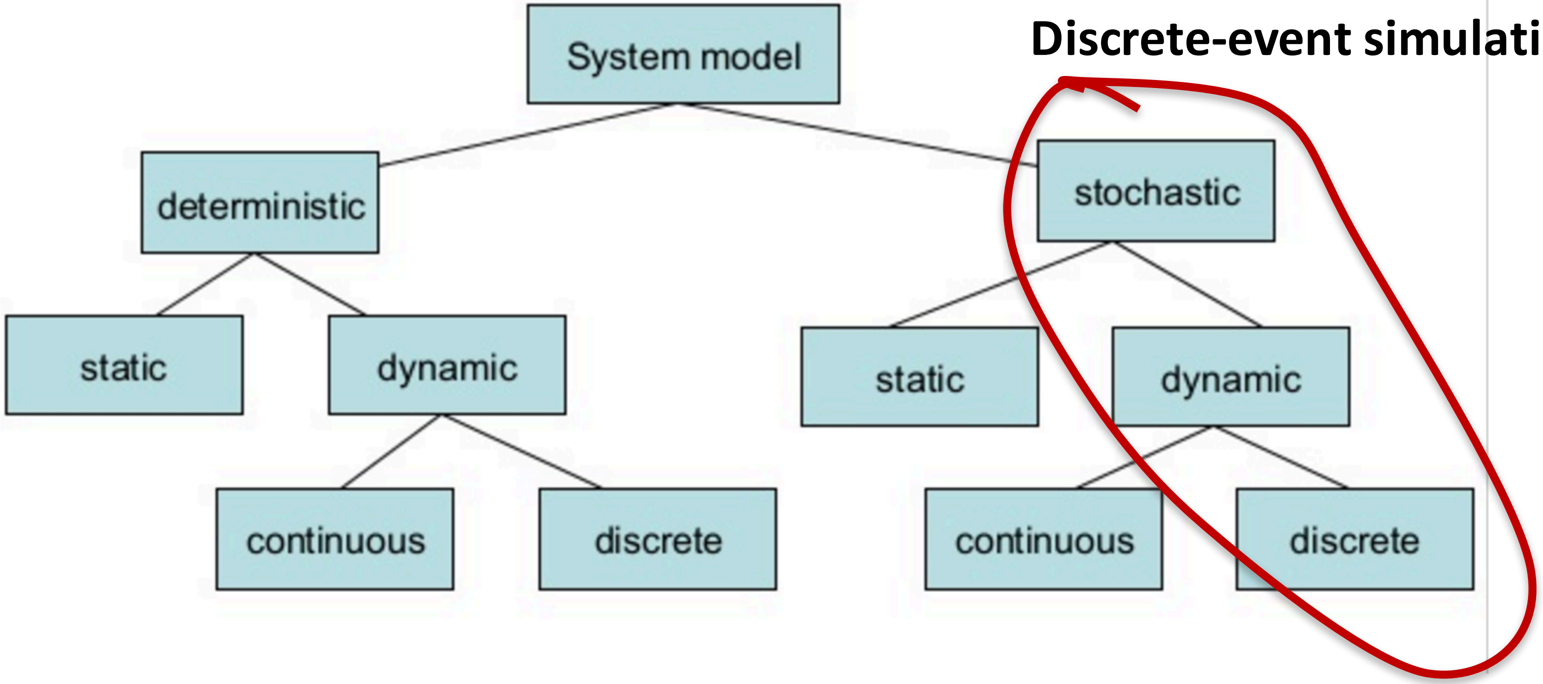


Fig. 9 Maximum repair time.

System Models

Discrete-event simulations



Our problem's model

Our problem is dynamic (time-dependent) and stochastic (all input isn't known, may be subject to change.), a stochastic and dynamic variant of the classical TSP.

Since uncertain data are gradually revealed, routes are not constructed beforehand. Instead, user requests are dispatched to vehicles in an on-going fashion as new data arrive.

Customer concept of the vehicle routing problem is failed elements in our case, and when they will need a repair (delivery when stocks are low in classic VRP) and their demand (time to repair) are unknown.
Moreover, the demand rate is usually low in the normal-mode of operation so that vehicles become idle from time to time. Relocating temporary idle vehicles is an issue.

- **Matisziw, T. C., Murray, A. T., & Grubesic, T. H. (2010). Strategic network restoration. *Networks and Spatial Economics, 10*(3).**
- **Duque, Pablo A. Maya, Irina S. Dolinskaya, and Kenneth Sörensen. "Network repair crew scheduling and routing for emergency relief distribution problem." European Journal of Operational Research 248.1 (2016): 272-285.**

Novelty

- Studies in the literature separately optimize (software)reprovisioning and (hardware)repair,
- do not consider a dynamic network environment where connections arrive/depart, new correlated or uncorrelated failures occur, and unpredicted delays in the repairs can happen.
- Reprovisioning considering service differentiation might change repair decisions greatly, hence jointly optimizing reprovisioning and recovery results in higher satisfied customers.

Uncertainties after a large-scale failure

- Type and number of failures
- Exact failure locations
- Severity of failures
- Road conditions

Hardware-level recovery

Recovery plans should be designed dynamic to be able to react properly to the unexpected conditions such as travel and repair times.

Software-level recovery

To alleviate the suboptimal scheduling (uncertainty in the operational environment): **software-level redistribution of resources based on service requirements.**

Software-level recovery

In a resource crunch scenario:

Exploiting network heterogeneity

- Degraded-service tolerance
- Importance (based on network operators' revenue) etc.

Based on service differentiation, reallocating the existing network resources among competing working and disrupted connections alleviates disruption problem until more resources become available through physical repairs.

Physical-level recovery

Determine schedule dynamically

i) making the repair crew/failure assignment based on the current network state (consider the recovery teams that are expected to be available soon)

ii) relocating the recovery teams to place them as close as possible to the most critical sections, which will give the highest damage in case of failure.

iii) better real-time assignment and routing decisions can be made if uncertain data estimations (derived from historical data) are used.

Problem: to make **as much as** connections operational **as soon as** possible exploiting service differentiation and deciding on the best repair plan considering reprovisioning (hence urgent remaining disrupted connections) with limited repair resources in a dynamic and realistic network environment where connections arrive, depart, new failure arrives, and expected repair plans change.

Service differentiation

Denied or deferred service: Deny service (given a time window, if not served, then reject) / **forward them to competitors** to avoid excessive delays/unacceptable costs.

Service differentiation is made based on importance (revenue), degraded-service tolerance , delay/down-time tolerance (a connection may tolerate advanced reservation or require immediate reservation. Some requires continuous connection while others allow some downtime as long as the deadline is met.)

Most of the connections allow some down time, even if it is very marginal. That does not imply that those connections are delay-tolerant. **Delay tolerant connections** are considered to be satisfied as long as the deadline is met. (backups and synchronization of data stored in multiple datacenters)

- **Availability**

Availability and Economics of 9s

- Availability is measured in terms of 9s—one to five nines
- An availability of five nines indicates that the application is available for 99.999% of the day. This translates to an uptime or availability of 86,399,136 milliseconds in a day that consists of 86,400,000 milliseconds.

Table 9-1. *Cloud Applications' Revenue*

Business	Revenue/Year (2013) USD	Revenue/Minute USD
EdiActivity	500,000	1.00
GXS	480,000,000	913
Salesforce	4,070,000,000	7743
eBay	16,050,000,000	30,536
Southwest Airlines	18,610,000,000	35,407
Google	59,730,000,000	113,641
Amazon	74,450,000,000	141,647

Hardening Azure Applications, pp 133-144, chapter: Availability and Economics of 9s by Suren Machiraju (Bill Gates foundation) and Suraj Gaurav (Microsoft), 2015.

Economics of non-availability

- Loss of reputation
- Customer and partner dissatisfaction
- Risk of regulatory oversight
- Loss of sales
- Lost and damaged data
- The need to restart in order to return to full operation
- Lowered employee morale
- Inconvenience, strife, accidents, loss of life, and human tragedies

on average, a single hour of downtime per year costs a business over \$100,000, while over 50% of businesses say the cost exceeds \$300,000 per minute, and one in 10 indicate that an hour of downtime costs their firms \$1 million or more annually. (survey of 600 orgs. 2014 ITIC)

$$\text{Availability} = \frac{(\text{Total Time}(-)\text{Downtime})}{\text{Total Time}} \times 100$$

Table 9-2. 9s, Uptime, and Downtime for a Total Time of 43,200 Minutes per Month

Availability Target	Minimum Uptime: minutes/month	Maximum Downtime: minutes/month
99.9999%	43200	0.0432
99.999%	43200	0.432
99.99%	43196	4.32
99.9%	43157	43.2
99%	42768	432
90%	38880	4320

For an availability target of 99% you are allowed 432 minutes, or about seven hours a week, of downtime; at 99.9% you get 3/4 of one hour per week. Yes, each 9 on the availability target does mean there was a significant reduction of your application's downtime.

It is very common to measure availability in monthly intervals.

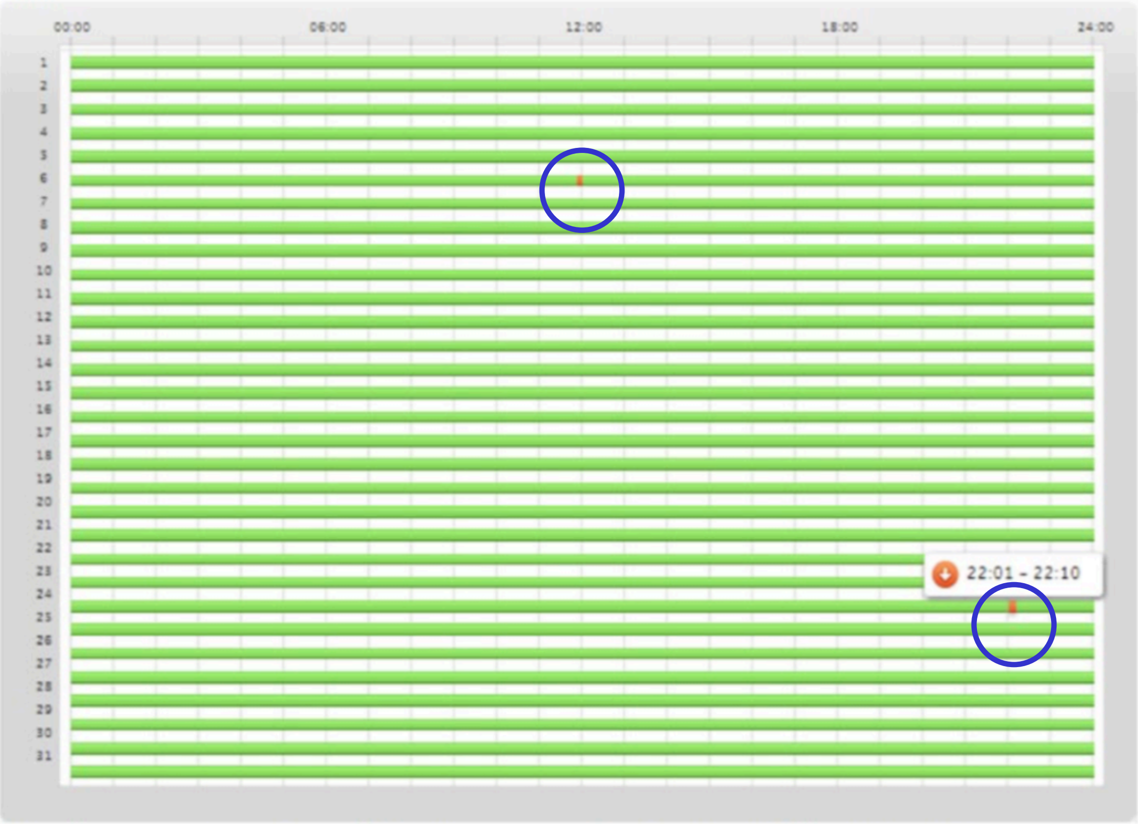
A commercially available service

GC EdiActivity (Jan 2015)

Recent | History

Last checked	Uptime this month	Avg. resp. time this month	Check type: HTTP
04/12/2015 10:41:29PM	99.96%	319 ms	Check resolution: 1 minutes

Uptime



Day-by-day availability (uptime) for the selected month. Red sections indicate downtime. Hover mouse pointer over sections to get exact times.

Outages

From	To	Duration ▾
01/06/2015 11:51:29AM	01/06/2015 11:59:29AM	8m
01/24/2015 10:01:29PM	01/24/2015 10:09:29PM	8m

Enforcing availability via SLA

(g) For Exchange Online Protection (EOP):

With respect to EOP licensed as a standalone Service, ECAL suite, or Exchange Enterprise CAL with Services, you may be eligible for Service Credits if we do not meet the Service Level described below for (1) Uptime and (2) Email Delivery.

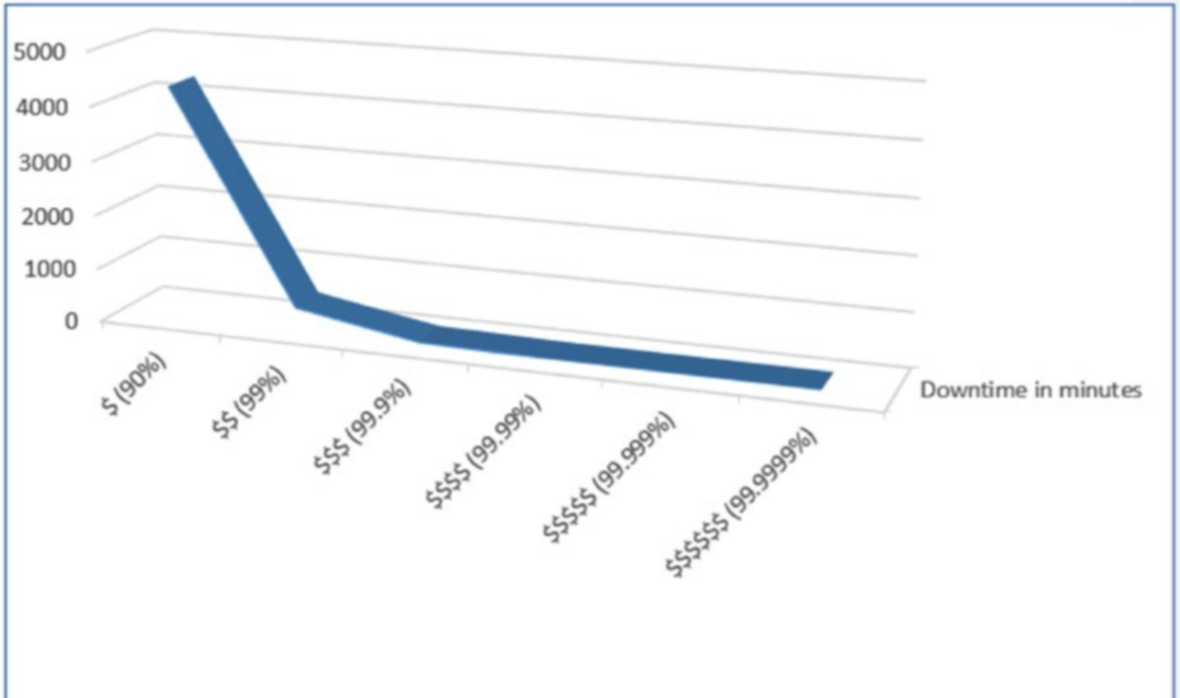
1. Monthly Uptime Percentage:

If the Monthly Uptime Percentage for EOP falls below 99.999% for any given month, you may be eligible for the following Service Credit:

Monthly Uptime Percentage	Service Credit
<99.999%	25%
<99.0%	50%
<98.0%	100%

Figure 9-3. *Service credits associated with SLAs*

This customer's availability requirement is straightforward: there is an SLA promising 99% availability during business hours. so, as long as ediActivity ensures preventative maintenance, and updates are done outside of the stated business hours, the application meets the SLA. Technically, the customer is seeking an availability slA of 37.5%, but only cares about business hours, so this is not a problem.



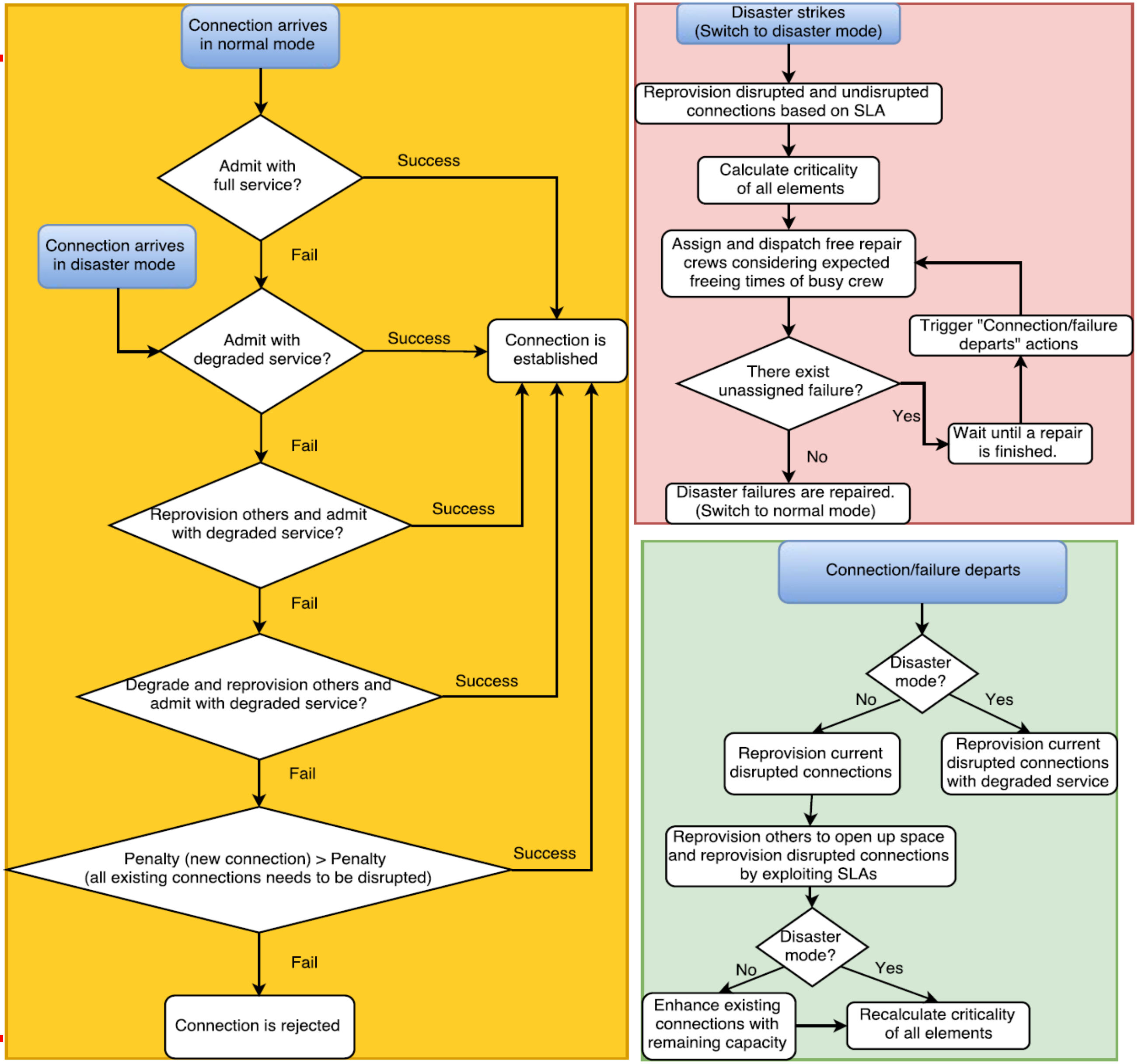
	$ (90%)	$$ (99%)	$$$ (99.9%)	$$$$ (99.99%)	$$$$$ (99.999%)	$$$$$$ (99.9999%)
Downtime in minutes	4320	432	43.2	4.32	0.432	0.0432

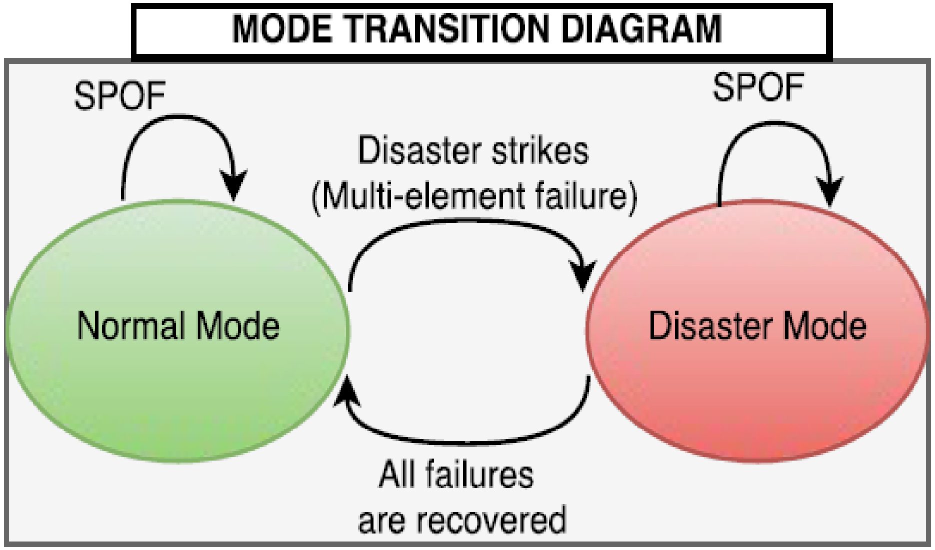
Figure 9-5. *Mapping availabilty to downtime*

Algorithm

- A **time horizon**, also known as a **planning horizon**, is a fixed point of time in the future at which point certain processes will be evaluated or assumed to end.
- Heuristics with some **look-ahead** capability should be developed.
- At any time each driver just needs to know his next stop. Hence, based on the revealed uncertainties, the schedule dynamically changes.
- Incorporate expected freeing times of the unavailable vehicles. If we direct vehicles without using the future information, solution will be very suboptimal.
- The value of repairing a link depends on the number of connections that can be routed after it is operational.
- Objective is throughput optimization, i.e., the maximization of the expected number of requests satisfied within a given period of time.

Dynamic recovery flow charts





Two modes of operation, where connections might have different requirements. Disaster mode is the mode of operation starting with the strike of a disaster and ending after repairing all effected elements and normal mode is the mode used rest of the time.

Disaster-mode of operation relaxes connection SLAs and may tolerate more down time

Possible extensions

- Dynamic relocation of idle recovery teams based on centrality of the links (calculated by the sum of connections' importance parameter it is carrying). This will help us to react faster to the failures (uncorrelated), which give the most damage. Here, we can investigate the areas recovery teams are responsible for, and how far we should send a repair team from its responsible zone.
- Improved estimation of vehicle travel times and repair times can be available as the repair crew gets closer to the damaged region. We do not need to wait certain information to make the final decision, as improved estimations arrive, dynamically update repair schedule. (near future -> gets more computational power vs distant future) (how much these micromanagements affect with sw recovery.)
- All teams are assigned to the current failures (from NY to CA?) / Responsibility zones?
- Joint optimization of road network and communication network.