Internet Services Need High Availability

xxxxxxxxxxxxx

Department of Computer Science, Rutgers University
110 Frelinghuysen Rd, Piscataway, NJ 08854

Abstract

People argue that five nines availability is unnecessary for Internet Services. We beleive this is not true. In this paper, we discussed why "five nines availability is unnecessary" is not true and provided some economic results to support our point. Finally, we briefly mentioned an active research area: how to improve Internet Services' availability.

1 Introduction

Although there is no clear definition of Internet Services, people tends to agree to use Internet Services to refer those giant-scale services, such as data-centers, search engines, internet directories, online shopping centers, etc. The common features of these services are that they have a large amount of customers and their customers are distributed across a wide area.

Availability is defined by many researchers as equation 1. MTTF is the mean time to failure. MTTR is the mean time to repair. People use the number of nines in the availability when referring to availability numbers. For example a 99.9% availability is three nines, while 99.999% is five nines. A three-nine available system means that it will have less than 525.6 minutes down-time per year, while a five-nine system will have less than 5.2 minutes down-time per year(table 1). A personal computer's availability is three nines.

$$Availability = \frac{MTTF}{MTTR + MTTF} \tag{1}$$

As more and more people depends on Internet Services, there is an argument about how available an Internet Service should be. Some people argue that Internet Services don't need five nines. We believe this is wrong. We believe that Internet Services need high availability, i.e., five nines availability is necessary.

Availability(%)	Class	Average Down-time(m/year)
90	1	52560
99	2	5256
99.9	3	525.6
99.99	4	52.5
99.999	5	5.2
99.9999	6	0.5

Table 1: Down-time of Each Class

The organization of this paper is as following. In section 2, we will discuss the argument Internet Services don't need five nines and why it's wrong. In section 3, we will present more arguments why

Internet Services need high availability. Finally, in section 4, we will discuss our conclusions.

2 Internet Services Don't Need Five Nines and Why It Is Wrong

People argue that Internet Services don't need five nines. Their reason is that the end user's availability is at most three nines. Most people wouldn't notice the difference if their service was at 5 nines. So, obviously, five nines is not required.

People have done a lot of work on the end user's availability. For example, Merzbacher and Patterson showed that the availability of Internet Services are mostly affected by the end user's availability[4].

Although the conclusion that the end user affects the Internet Services availability mostly is correct, we still can't make the conclusion that the end user won't notice the difference of Internet Services' availability and five nines is unnecessary for Internet Services. The reason is that there are some hiding assumptions behind this argument, which is wrong, or at least infeasible:

1. *Continuity* PC functions continuously. In their experiment, the end user's system functions continuously, which is not always the case for most end user.
2. *Dependence* When an Internet Service fails, the user's PC will also fail, so the user won't notice the Internet Service's failure.
3. *Indistinguishability* Users can't distinguish an Internet Service outage with a local outage and they treat and tolerate them as the same problem.
4. *Single User* To make the conclusion that Internet Service's unavailability wouldn't be noticed by the end user, another assumption is hidden: there is only one or very few users in one Internet Service. Obviously this is not the case.

While these assumptions can be easily identified as wrong, it's easy to be confused without a clear understanding.

2.1 How PC's Discontinuity Affects its Availability

One way that will affect a system's availability is whether it functions continuously. Discontinuity has two effects on a system's availability: rejuvenation and shorter working time.

The theory of rejuvenation says that a system will work better after it restarts. Even if we don't consider the effect of rejuvenation, a discontinuous system simply has shorter working time and makes it less probability to crash, which makes it has a higher availability than the same system which functions continuously.

The availability of a discontinuous system can be calculated by equation 2. p_t is the percentage the system functions. p_a is the system's availability when it functions continuously.

$$AvailabilityofaDiscontinuousSystem = 1 - p_t(1 - p_a) \tag{2}$$

From this equation, we can see that a system with three nines can achieve four nines if it only functions about 15% of the time. As personal computer users often shutdown their computers and only turn them on when necessary, we believe that personal computers only work 30-10% of the time. This makes their availability will be much higher than the estimated availability.

2.2 How Independence affects the user's observability

It's safe to assume that the Internet Service and the end system will fail independently, which means that the unavailable of an Internet Service will be noticed by an end user for sure.

Even if they do happen at the same time, for a three nines end system, the user still has 99.9% chance to see the failure.

2.3 Why Indistinguishability is Wrong

Another assumption of the argument is that the end user can tolerate Internet Services' outage just as his end system's outage. This is easily to be identified as wrong.

The end system's outage is often scheduled, predictable and known to the end user. While the user checks his local environment, it's easy to notice the outage of the Internet Service. Obviously, a unscheduled unknown remote system outage is more annoying and intolerable than a known scheduled local system outage.

2.4 How Multi-user Affects the Availability Requirement

The expectation that an Internet Service outage won't be noticed by an end user is impossible. For a three nines end system, suppose their failure is independent. If there are two such systems, the possibility that there is always one is up is $1-(1-p)*(1-p)$. For a three-nine end system, the probability that there will always be one up is six nines. The expectation of the number of up systems is $p*(1-p)+2*p^2$=1.996, which means that there are at least one is up. With millions hits a day for an Internet Service, the outage of an Internet Service will easily be observed by thousands of users.

2.5 Summary

All the above shows that although the end system's availability does affect the end user's availability, as a whole, Internet Services need high availability to better serve its users.

3 Internet Services Need High Availability and How to Improve the Availability

In this section, we will discuss some commercial arguments about why Internet Services need high availability. The reason we are making commercial arguments is that today's Internet Services are all running by some company. One of their objectives is to make money. We believe that a high available Internet Service is not only interested to the academic researchers, but also interested to these companies. The reasons are:

1. *Customer Interest* The customers will favor high availability.
2. *Outage Cost* The cost of an Internet Service is very high and is worth the effort to have a high Internet Service availability.

3.1 Customer Interest Requires High Availability

Researches have shown that unavailability will affect a user's interest and will make the service lose customers[1]. When a user is ready to do something and the service of a site is not available, often the user will seek an alternative site. This means that the site will lose a potential customer. With today's companies trying so hard to attract a visitor, losing a customer due to unavailability is intolerable.

3.2 Outage Cost

Although losing a potential customer may not mean too much to a big company, the cost number for some real outages may mean something.

Researchers have done a lot of work on the outage cost of an Internet Service. Table 2 is the result of Paterson[6].

Brokerage operations	$6,450,000
Credit card authorization	$2,600,000
Ebay	$225,000
Amazon.com	$180,000
Package shipping services	$150,000
Home shopping channel	$113,000
Catalog sales center	$90,000
Airline reservation center	$89,000
Cellular service activation	$41,000
On-line network fees	$25,000
ATM service fees	$14,000

Table 2: Cost of one hour of down-time

From table 1, we can see that an Internet Service with three nines will have about 10 hours down-time. Combined with table 2, we can see this will cost them millions of dollars. This will be too expensive. With so high cost, it's worth to improve the Internet Service's availability.

4 Conclusion

In this paper, we argued that Internet Services need high availability. We pointed out that although the end system's availability affects the end user's experience mostly, we believe that Internet Services still need high availability to serve the large number of users.

After we argued that Internet Services need five nines, the next thing will be to improve today's Internet Service's availability. With today's technology achievement, we believe that it's possible to achieve five nines availability. A great deal of work has been done on how to improve system and internet service's availability[5, 2, 3], etc.

Researchers have pointed out that software and operator mistakes are the main source of system outage. We believe that ideas proposed in [5, 3] will be helpful to improve availability: partition, testing, configuration, periodically and partially restarting, etc. This is an active research area, and we'd expect to see great result in the near future.

References

[1] R.H.Frankand B.S.Bernanke. Principles of economics. pages 92–3, 2001.

[2] J. Gray. Why do computers stop and what can be done about it? In *Proceedings Fifth Symposium on Reliability in Distributed Software and Database Systems*, Jan. 1986.

[3] Taliver Heath, Richard P. Martin, and Thu D. Nguyen. Improving cluster availability using workstation validation. In *SIGMETRICS 2002*, June 2002.

[4] M. Merzbacher and Dan Patterson. Measuring end-user availability on the web: Practical experience. In *International Performance and Dependability Symposium (DSN)*, pages 184–194, June 2002.

[5] David Oppenheimer, Archana Ganapathi, and David A. Patterson. Why do internet services fail, and what can be done about it? In *USITS 2003*.

[6] David A. Patterson. A simple way to estimate the cost of downtime. In *LISA 2002*, April 2002.