Essay Assignment 2 EE 4th grade 李彥融 B98901180

BitTorrent File Distribution and Epidemic Diseases

Introduction

I always use uTorrent to download videos such as latest Japanese animes. I found p2p is not only convenient but really fast. Usually it takes only a couple of minutes to reach maximum download speed, and then finishes in another few minutes downloading the whole file, which is over 100 mb. I think the reason is that the torrent is pretty new, only hours young, and hundreds to thousands of people are downloading in the same time. However, it's not always lucky. Once I found a torrent file released just 10 minutes ago. I got the torrent and started to download, but it didn't have any speed until half even an hour later. Start downloading as early as possible doesn't mean you'll finish early. It seems like if you wait for a while, you'll download in the most efficient way, spending least total time downloading.

It reminds me of epidemic diseases. The probability of getting infected is low at the beginning because the number of infected people is small. Then the probability rises day by day, reach a peak at certain time, then falls to zero after everyone is recovered and becomes immune. In this essay I'll try to use epidemic models on BitTorrent file distribution, decide whether it is valid, and use it to predict when to start download will be better.

Epidemic model: the SIR model[3]

A simple model describing epidemic diseases transmission is SIR model. SIR stands for the susceptible, the infected, and the recovered. The model is,

$$N = S + I + R, \qquad \frac{dS}{dt} = -\beta S I, \qquad \frac{dI}{dt} = \beta S I - \gamma I, \qquad \frac{dR}{dt} = \gamma I.$$

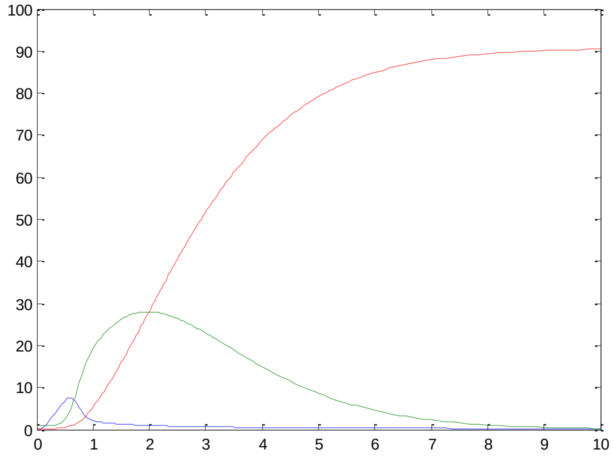
The total number of people N is assumed to be constant. In analog to BitTorrent, *S* are peers still downloading; *I* are those finishing download and still seeding; *R* are those stop seeding and leave the network. The big pattern quite matches, unless a key part missed, the new coming peers. In the epidemic model there is a natural birth term, I adjust it to this form,

$$\frac{dS}{dt} = -\beta S I + \mu\, t\, exp\left(-\frac{t}{\tau}\right)$$

I assume coming rate is of t*exp(t) form. Use MATLAB and the following input commands,

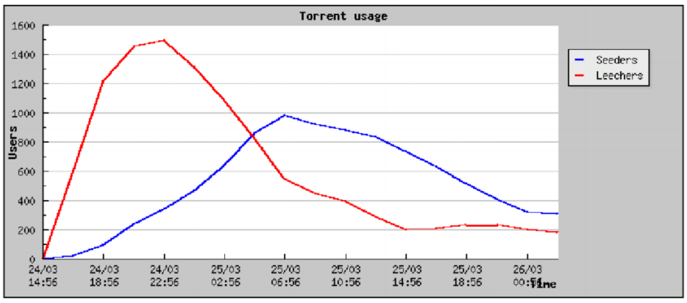
```
>> f1 = @(t,y)[-1*y(1)*y(2)+100*exp(-t)*t; y(1)*y(2)-1*y(2); 0.9*y(2)]
>> [T,Y] = ode45(f1,[0 10],[[0],[1],[0]])
>> plot(T,Y)
```

We can get the plot with [blue, green, red] curves are [y(1), y(2), y(3)] or [S, I, R].



Result Discussion

The above is just a rough result of what might be the case in BitTorrent. The units are not well considered. Though the pattern varies with parameters, there is still something can be discussed.



The trends seem quite correct, at least in terms of my experience and a real measurement found on the Internet (right up figure[4]). Number of downloaders excesses seeds at the beginning. Later, number of seeds reaches maximum, where I have a higher chance to complete download with the highest download speed. Finally all the people finish download and leave network, and the late comers will find it hard to download because the seeds are too few.

To be improved

Though the big picture seems correct, problems indeed exit. In the above simulation, I assume the probabilities of a downloader to reach every seed are equal. I assume a downloader only have two states, no speed and highest speed. When enters highest speed, download time is short and can be neglect. So a peer change from downloader to seed immediately, like a uninfected people transits to infected. This assumption is based on my experience. The peers I connect are usually fixed when I am in highest download speed, which is analogous to 1 to 1 infection model in epidemic model (I'm not sure whether it is 1 to 1 in epidemic, though).

When I dig into the specification of BitTorrent protocol, I found I overlooked a lot mechanism like tracker server, interested and choked flags, etc. Hope it will not affect the results too much.

Conclusion

Though there are still much to be improved, the model based on epidemic diseases can be used to describe the behavior of BitTorrent downloaders, such as not to start download too early would be helpful in efficiency (though I didn't give a certain time, left to be done). The rest is to do more measurements, refine the model, and to verify whether this approach is correct, or a useless result which is just derived from what is already known.

References

1. BitTorrent, Wikipeida, http://en.wikipedia.org/wiki/BitTorrent
2. BitTorrentSpecification, TheoryOrg, http://wiki.theory.org/BitTorrentSpecification
3. Epidemic model, Wikipedia, http://en.wikipedia.org/wiki/Epidemic_model
4. http://www.cs.toronto.edu/~walex/mpvc/UnderstandingBitTorrent.ppt