1. If an infected computer randomly picks a computer in the address space that's already infected, what should the behavior be?  Should it basically ignore that computer and move on to the next scan?

Answer: if a scanned target is an already infected computer, nothing will happen. The scan is treated as if it missed a target and the computer should continue its scanning process.

1. I'm a little confused about how to implement the 2nd part of the assignment.  If the worm infects computer A at time = 10, then computer A goes into "being infected" and I then calculate X using Geometric distr and find out that X = 3.12 for instance.

Answer: First, Geometric distribution always generates integer random variable, i.e., X is always an integer. Second, if the worm infects computer A at time 10 and X is chosen to be 3, then computer A will begin to send out scans to infect others at time 13.

1. I wasn't sure what the best way to generate a random integer.

Answer: If you program in C and if you use the rand01() function I provided in my lecture notes “generate-rv”, then you can use:

Address = int( floor(rand01() \*Max) );

This will generate an integer value between 0 and Max-1.

1. To return X, the number of time units it takes to simulate traffic delays, I would first generate a geometric distribution, and then randomly generate a probability, look up that probability in the geometric distribution and then return the X number of time units.  Is that correct?

Answer: You generate a (0,1) uniform distr. value, then following the example in lecture notes “generate-rv”, page 11 (or page 3 in notes “statistic-matlab”), to obtain the value of X. You should have known this after finishing project 1.

1. The probability to hit a target is very very low. In the project, the network space has 2^14=16384 IP addresses, but the network only has N=400 vulnerable computers. Thus the probability for one scan hitting a potential target is 400/16384 = 0.024, i.e., on average a worm infected computer needs to send out 16384/400 = 40.96 scans before it hits one target. Even worse is that some hit targets that have already been infected.
2. How should I represent computer’s IP address?

Answer: You do not really need to use “128.119.0.5” to represent an IP address, you can use an array and use the array index to represent the addresses in the network IP space.

1. Remember that each scan has the probability of 400/2^14=0.025 to hit one of those 400 machines. That means on average 40 scans generate one hit. Thus the one initially infected machine should take on average 20 unit times to cause one infection (i.e, I(t)=2 when t is around 20).

1) Since every infected machine will scan until it finds another vulnerable  
> machine, is it safe to say that the infected machines at any time will be in  
> the order of  1,3,6,12,24,48,96,192...........398.

No. a scan is targeting a randomly selected IP address, so which  
vulnerable computer is infected first is totally random. In addition, not all scans can cause infection. In fact, most scans will be wasted since the number of vulnerable computers is very small compared with the total IP space scanned by the worm.

> 2) If an infected machine will scan until it fnds a vulnerable machine, will  
> the time icrease as it goes through the scan each time, or will it wait  
> until after it infect a vulnerable machine.

This is a discrete time simulation, so at each discrete time tick,  
each infected machine only generates 2 scans. That is to say, at each  
discrete time tick, the program goes over all infected machines once,  
and let each infected generate 2 scans.

> 3) If (for example) ten machines are infected , the time should not change  
> until all ten machines infects vulnerable machine, Is that correct?

That's wrong. see my previous answer.

> 4) For the 100 simulations, are the plots suppose to be overlapping, if so  
> how would you find the average?

You only plot one curve for the average I(t) over 100 simulation runs.  
Do not plot 100 curves. I explained how to compute average I(t) for  
any given t in class. See my lecture video.

> 5) Should the 100 simulations plot overlap with the analytical plot, or  
> should it be the average of the 100 simulations that overlaps with the  
> analytical

The average I(t) curve from simulation should be a little slower than  
analytical curve.