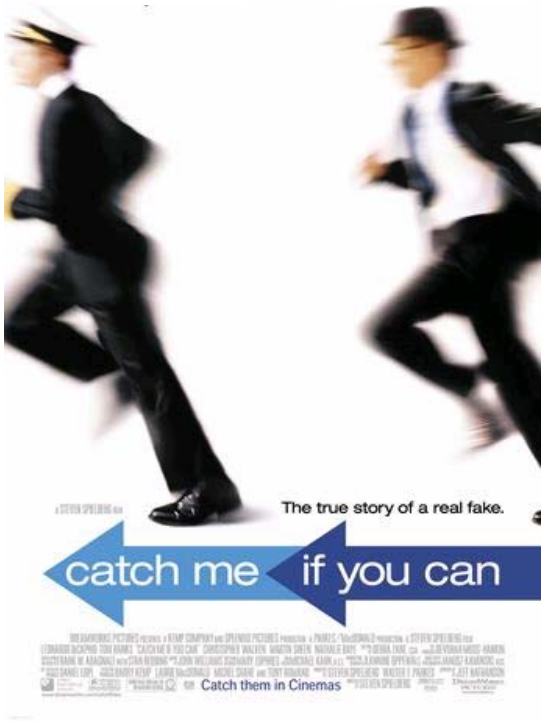
Catch me if you can – Anonymous broadcasting in Delay Tolerant Networks

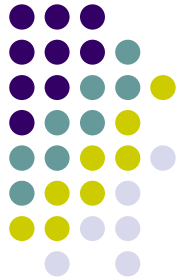
Xiaofeng Lu Pan Hui Crowcroft Jon



Why we need anonymity?

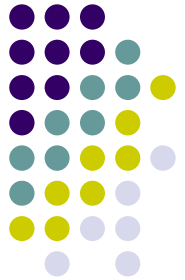
- In some untrust or hostile environment, we want to send some messages to others but not wanting others know the initiator of this message for the security reason, e.g. censorship from government.

Definition of Anonymity



- Anonymity is the state of being not identifiable within a set of subjects, the anonymity set.

Related work



- People have already studied anonymity on Internet.(Crowds, Onion Routing)

- Anonymity on mobile Ad hoc networks has also been studied.

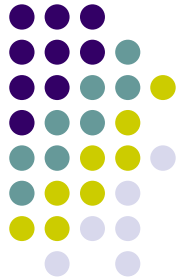
  Marie Elisabeth et al., Yanchao Zhang et al.

Question

- All these researches assume that the initiator needs routing protocol to cooperate with other people or nodes.

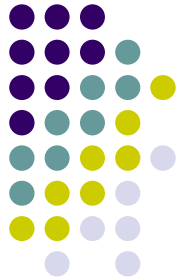
- Question: In a strange networks, if other nodes don’t have these protocols, could we still achieve anonymity?

Our study



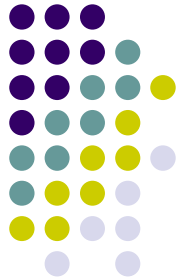
- We consider the scenario in which we send messages not to a specified node but as many as possible unspecified nodes – broadcasting.
- We don’t need some pre-installed protocols. The whole network don’t need to agree on specific protocols.

Adversary model



- We employ External Passive Local adversary model.
- An external adversary captures communication between nodes
- A passive adversary only eavesdrops the communication and does not modify it.
- A local adversary only control part of the network. No one know the whole network.

Some spy are aroud you



- There are some nodes which we called compromised nodes controlled by the authority. These compromised nodes are normal devices that can only receive messages within their communication ranges.
- When they receive some unwanted messages, they try to report their records to the authority. The authority find the location of the initiators.

Delay Tolerant Network

- We assume people use laptop or PDA to receive or send messages, and because people usually carry these devices with them, the networks consist of these devices is mobile social networks.
- As people would move from here to there, sometimes some nodes can talk with other nodes, sometimes they couldn’t, the network is Delay Tolerant Network

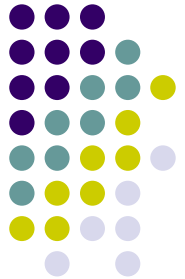
Secure enough?

- Imagine we want to send a “message” , and there are three compromised nodes within our node’s communication area.
- These three compromised nodes will receive our message at different time, because distances from these three nodes to the sender are different.

Authority Server: I can find you

- These compromised nodes then report their records to an authority server.
- The authority server employs the three records to find our node's location *p* using triangle localization algorithm.
- The smaller the node's distance to the location *p,* the larger probability the node to be the initiator.

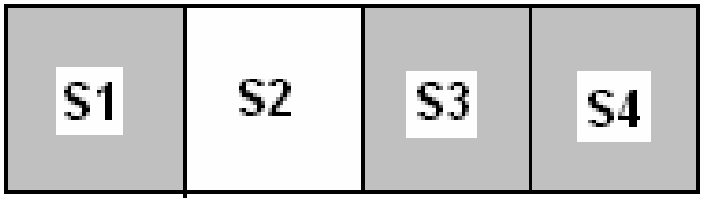
Split the message



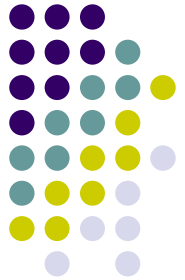
- We split a message into many seriate segments, seg 1, seg 2,…and send these segments at different time.
- Because the network is mobile social network, these segments will be received by different neighbors at different time.

You don’t know what I said

- In the beginning, these receivers usually can just receive parts of these segments, through which they could not know the content of this message.

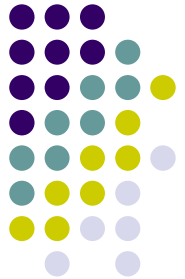


Game theory



- The receivers including compromised nodes have to help the initiator to relay these segments.
- Firstly, they don’t know the content of the message.
- Secondly, they also need others to help them to relay their messages. If it does not help others, no one would reply their message also.

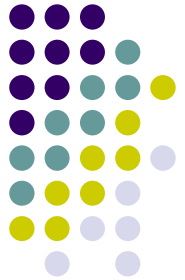
Receive all segments



- After some time, a node would receive all these segments, then they can read the content of this message.

- But, what these segments that the compromised node received would look like? Assume a node has a buffer to save all messages it received.

A example



- Assume the initiator splits the original message into 3 segments, s1,s2,s3.

- The compromised node would receive these segments from different nodes at different time. The format of their records should be (id of segment, arrival time)

- compromised node 1
- (s1,t1) (s1,t2) (s1 t3) (s2,t4) (s1, t5) (s2,t6) (s3,t7) (s2, t8) (s1,t9) (s3, t10)

- compromised node 2
- (s1,t2) (s2 t3) (s1,t5) (s3,t6) (s2, t8) (s1,t9)

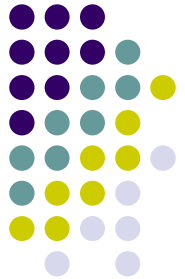
- compromised node 3
- (s2,t4) (s1, t5) (s3,t6) (s2, t8) (s1,t9) (s3, t10)

Which record it would employ?

- Assume the compromised node only uses the earliest time a segment arrives. These records would become:

- compromised node 1: (s1,t1) (s2,t4) (s3,t7)
- compromised node 2: (s1,t2) (s2 t3) (s3,t6)
- compromised node 3: (s1,t5) (s2,t4) (s3,t6)

A wrong location



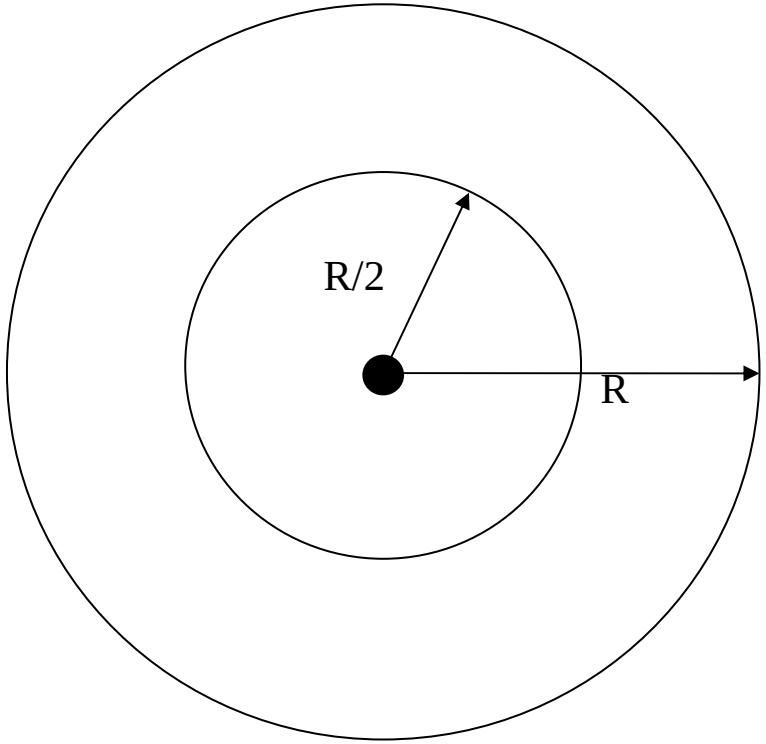
- If the authority server employs the difference of the arrival time to calculate the initiator’s location, obviously it would get a wrong location.

When should I send a message

- In what kind of condition, the initiator can send a message with higher anonymity?

- Because the initiator don’t know how many neighbor there are around it, it will send a Test Message before it sends its real message. Assume it receives k message relayed by its neighbors.

- What is the probability of all these replyers being within the small circle region?



- A: the region of the big circle, radius r
- B: the region of the small circle, radius r/2
- N(A): the number of nodes in area A
- N(B): the number of nodes in area B
- The probability of all these replyer being within area B is

$$\Pr(N(B)=k \mid N(A)=k)$$

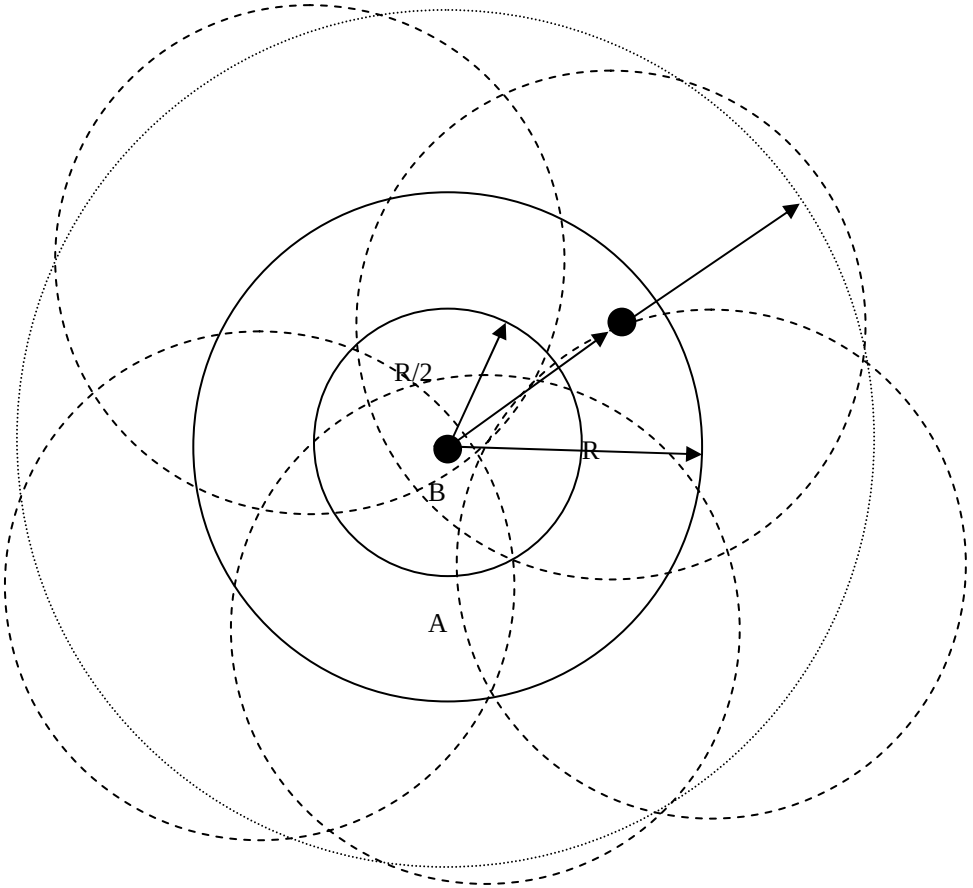
K is the number of received being replied messages

- According to Spatial Poisson Processes,

$$P(N(B) = k \mid N(A) = n) = \binom{n}{k} \left(\frac{|B|}{|A|}\right)^{k} \left(1 - \frac{|B|}{|A|}\right)^{n-k}$$

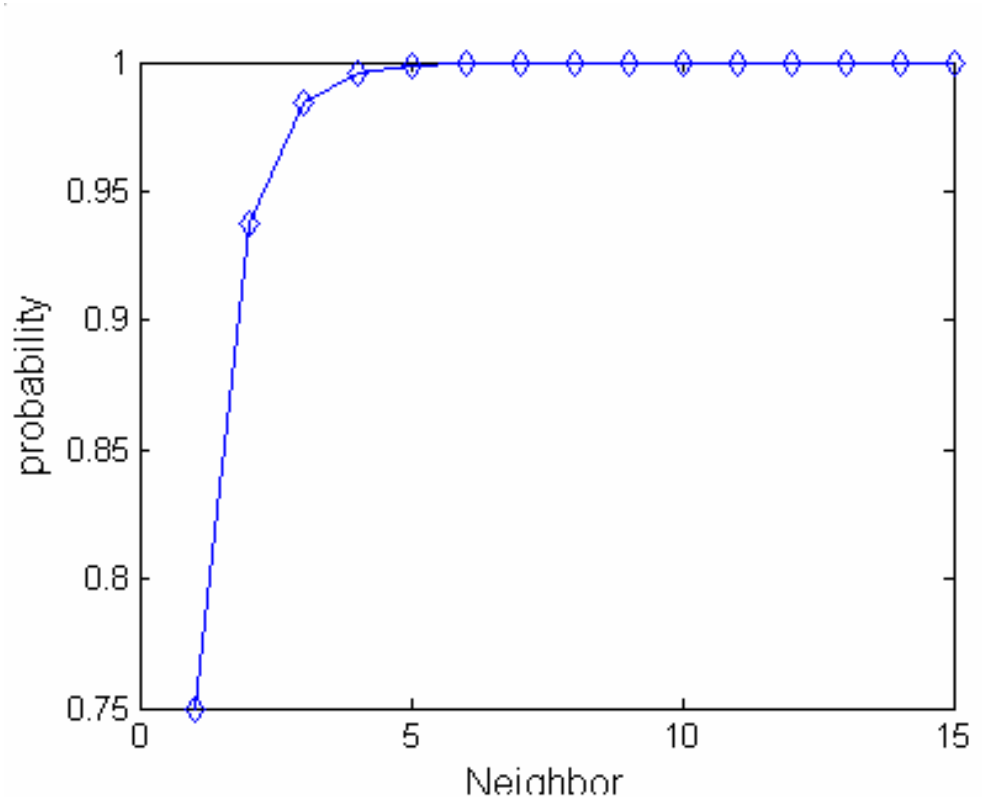
$$\Pr(N(B)=k \mid N(A)=k) = \left(\frac{|B|}{|A|}\right)^{k} = \frac{1}{4^{k}}$$

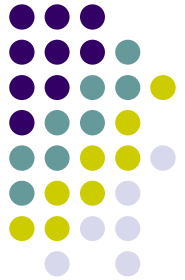
- If there are some nodes in the area A-B, then the message has a probability to spread out (r/2,r)



The probability of spreading

$$\Pr(\exists x, x \in A - B \mid N(A) = k) = 1 - \left(\frac{|B|}{|A|}\right)^k = 1 - \frac{1}{4^k}$$

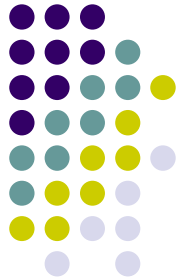




- This mean if the initiator can receive 4 or 5 reply messages, it can make sure the message could spread out.

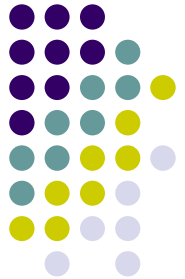
Quantify the anonymity

- Information theory has proven to be a useful tool to measure the amount of information.
- We try to measure the information obtained by the attacker or compromised node.
- Shannon’s definition of entropy allows to quantify the degree of anonymity of an electronic system.



- The adversary may assign some probabilities p_i to each sender as being the originator of a message, based on the information the adversary learned from the networks.

The entropy of the system

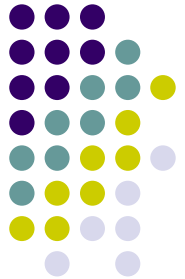


- We denote by $H(X)$ the entropy of the system after the attack has taken place.

$$H(X) = -\sum_{i=1}^{N} p_i \log_2(p_i)$$

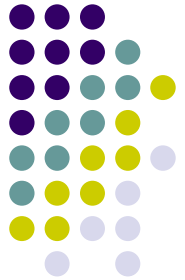
$$H_{\max}(X) = \log_2(N)$$

The degree of anonymity

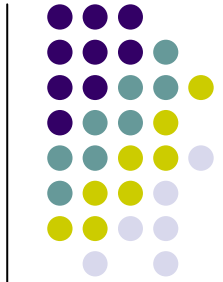


- The degree of anonymity is defined as

$$d = 1 - \frac{H_{\max} - H(X)}{H_{\max}} = \frac{H(X)}{H_{\max}}$$



- Let the suspicious area to be S, and the whole area to be W.
- Number of nodes in S is N(S).
- As we do not employ other technique to make the initiator more unsuspicious, all these nodes in S have the same probability to be the initiator.



$$H(X) = -\sum_{i=1}^{N} p_i \log_2(p_i)$$

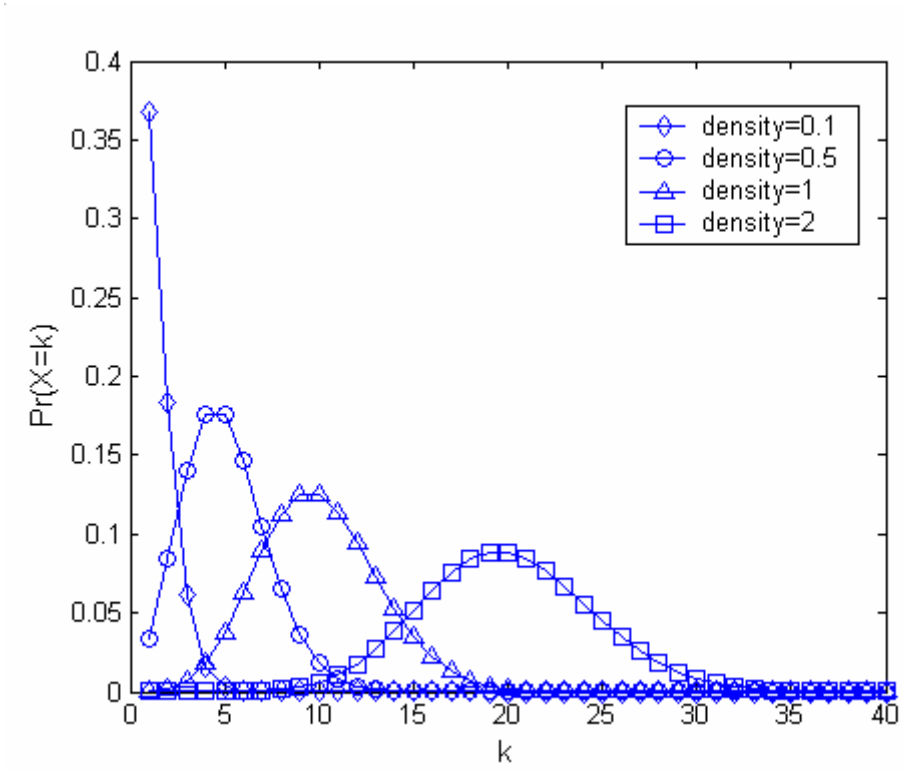
$$= -N(A)\left(\frac{1}{N(A)} \cdot \log_2\left(\frac{1}{N(A)}\right)\right)$$

$$= \log_2(N(A))$$

$$d = \frac{\log_2(N(A))}{\log_2(N(W))}$$

Spatial Poisson Processes

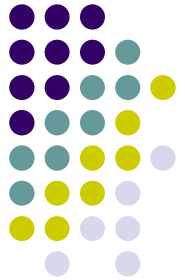
$$P(N(A) = k) = \frac{e^{-\lambda|A|}(\lambda|A|)^k}{k!}$$



- Select kj that has the largest probability

$$\Pr(\mathrm{N}(\mathrm{A}) = \mathrm{k}_{\mathrm{j}}) \geq \Pr(\mathrm{N}(\mathrm{A}) = \mathrm{k}_{\mathrm{i}}) \quad (\mathrm{i} \neq \mathrm{j})$$

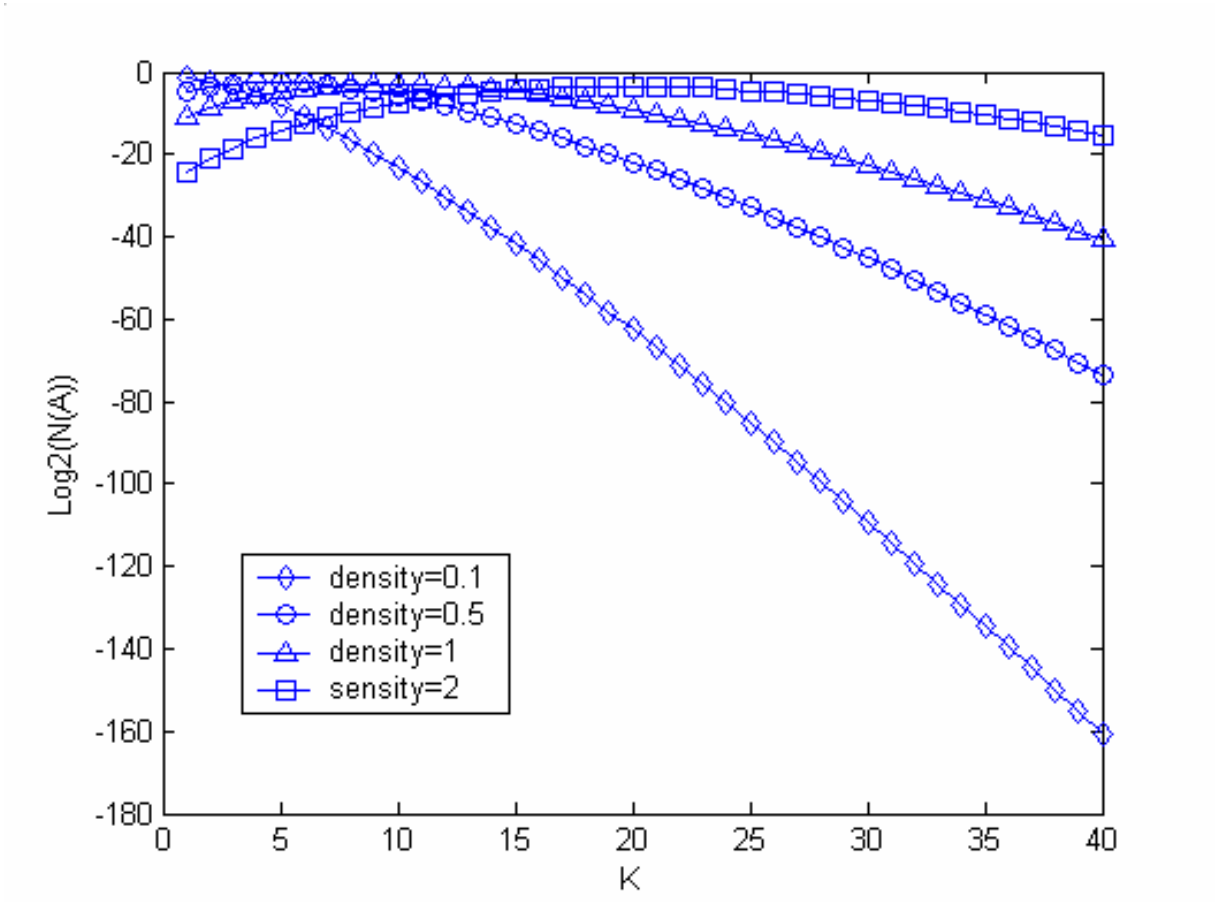
$$d = \frac{\log_2(Kj)}{\log_2(N(W))}$$



Thanks a lot!

Log2(N(A))

The variation of Log2(N(A)) with different node densities



Log10(log2(N(A)))

The variation of Log2(N(A)) with different node densities, the scale of Y axis is log10