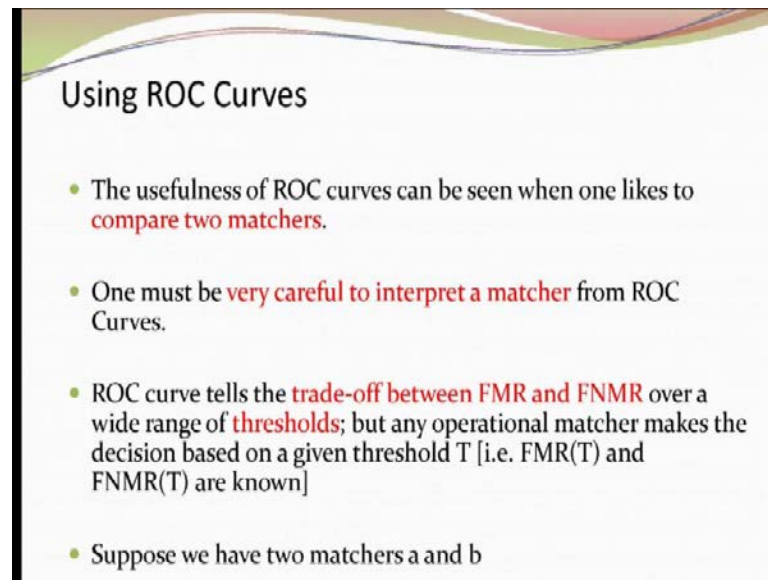


Biometrics
Prof. Phalguni Gupta
Department of Computer Science and Engineering
Indian Institute of Technology, Kanpur

Lecture No. # 13

(Refer Slide Time: 00:16)



Using ROC Curves

- The usefulness of ROC curves can be seen when one likes to **compare two matchers**.
- One must be **very careful to interpret a matcher** from ROC Curves.
- ROC curve tells the **trade-off between FMR and FNMR** over a wide range of **thresholds**; but any operational matcher makes the decision based on a given threshold T [i.e. $FMR(T)$ and $FNMR(T)$ are known]
- Suppose we have two matchers a and b

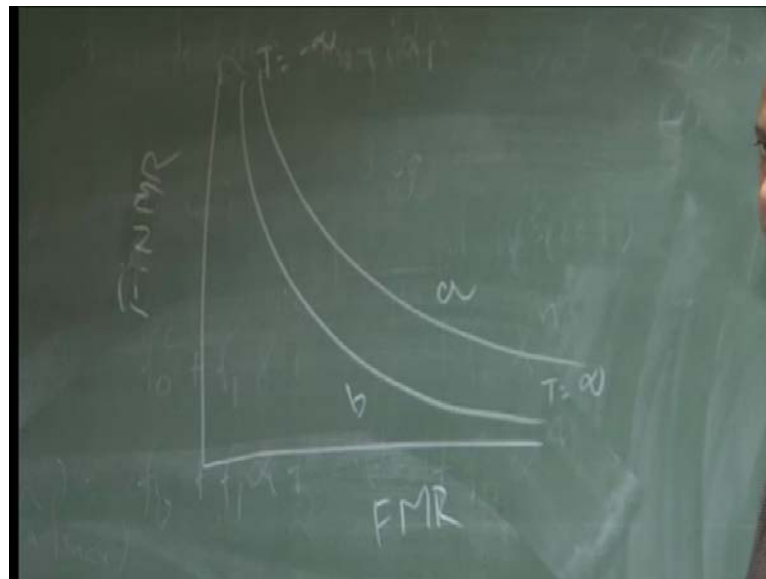
So, in the last class, we were discussing about the receiver, receiver operating characteristic curves; and also we mentioned the different ways, you can draw the ROC curves; and ROC curve gives you what the information you are getting in one side else no match rate, another side it will be false match rate, and corresponding threshold value that is the idea.

Now, it is not necessary that you have to use in one side exactly value of FMR and exactly value of FNMR. What we do? Sometimes depending upon the situation based on the number of people or number of participants in that database, you may have to go for log function, so that it becomes better way to represent the data. Now, these ROC curves useful, not only to see that FAR and FRR, sometimes what we do? We use it to compare the two systems. Suppose you develop a system, and obviously to justify your system performance, you will be telling that my system has the accuracy of this or you will be telling that FAR is this and corresponding FRR is this.

Now telling FAR and FRR, it varies upon different threshold value that for a particular threshold value, you will get one type of FAR and FRR for the different, you will be getting the different one. But your aim is to obtain the such a threshold value for which you are achieving the minimum FAR and also moderately good FRR or minimum FRR and moderately good FAR, you cannot expect both of them and will be minimum as I mentioned in one class that they are inversely related.

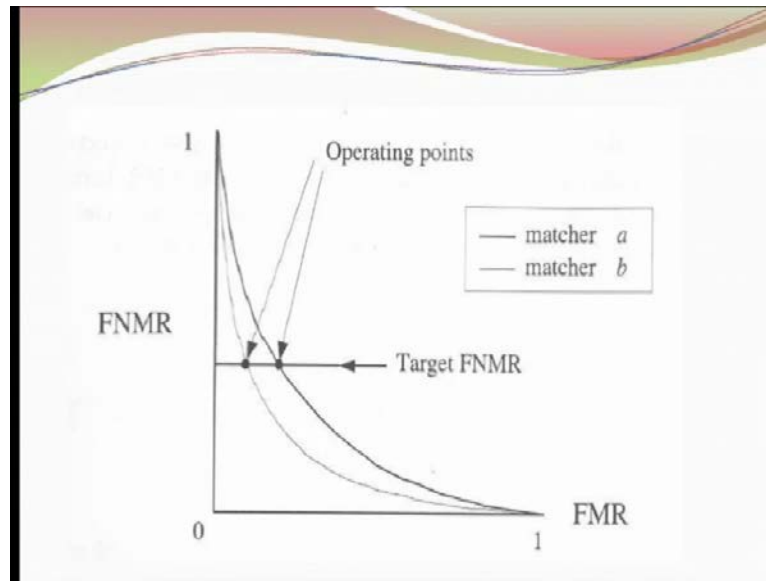
Now, this will be used to match compare the two matchers, but while you are comparing, you should be careful, because it is not directly you can tell, this matcher is better than the other one. I will come to that conclusion later on, but ROC curve tells you the trade of relationship between FAR and FRR.

(Refer Slide Time: 02:52)



As I told you there one is increasing, the graph is like this. And this ROC curves tells you that it is the trade of relationship between FMR and **FR** FNMR over a wide range of thresholds. So, if you fix a threshold then corresponding value of these two will be known or if you fix this one, the corresponding value of T and other part will be known. Now, suppose I have two matchers, which give whose RA ROC curve is like this, suppose this is b matcher and this is matcher a.

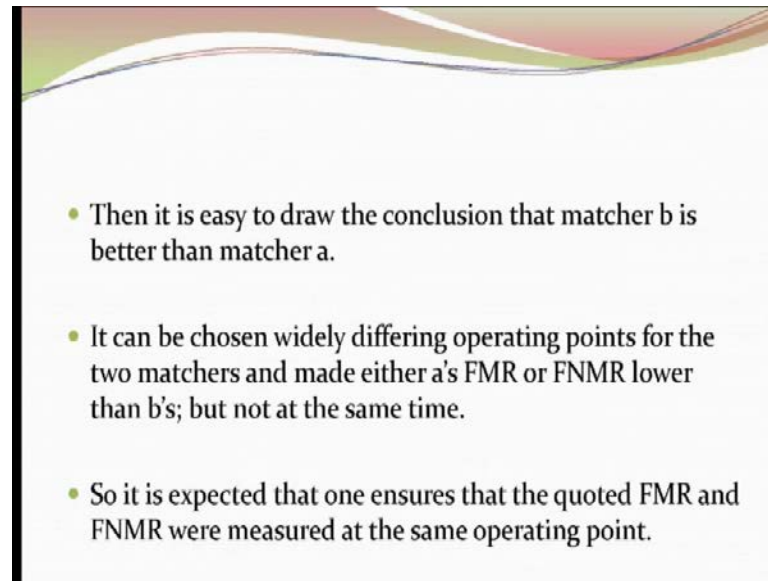
(Refer Slide Time: 03:53)



So, you got that two ROC curve; one is for matcher b, another one is matcher a. And if I tell you that which matcher gives you the good result. Obviously, you will be telling that this is giving you the better result, because this matcher gives you for any threshold value, this gives you that FMR or FNMR, one of them will be less than the other one.

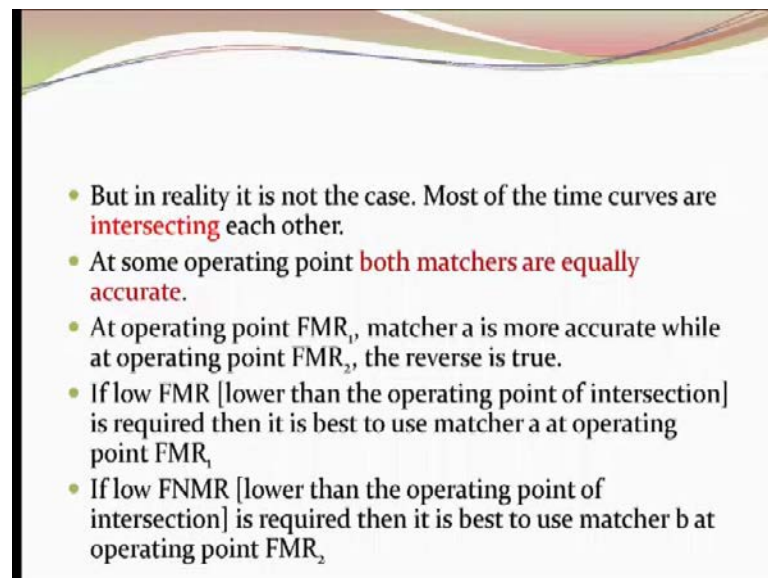
For this case, if I consider the same **FN** FNMR then for matcher b, FMR is less because this is the matcher b's FMR, and this is the FMR for matcher a. So, you can straight forward thing that you will be telling them matcher b is better than matcher a. And you take for any operating point or any threshold value or at any point, every point you will find the matcher b is better than matcher a, and you can draw the conclusion.

(Refer Slide Time: 05:02)



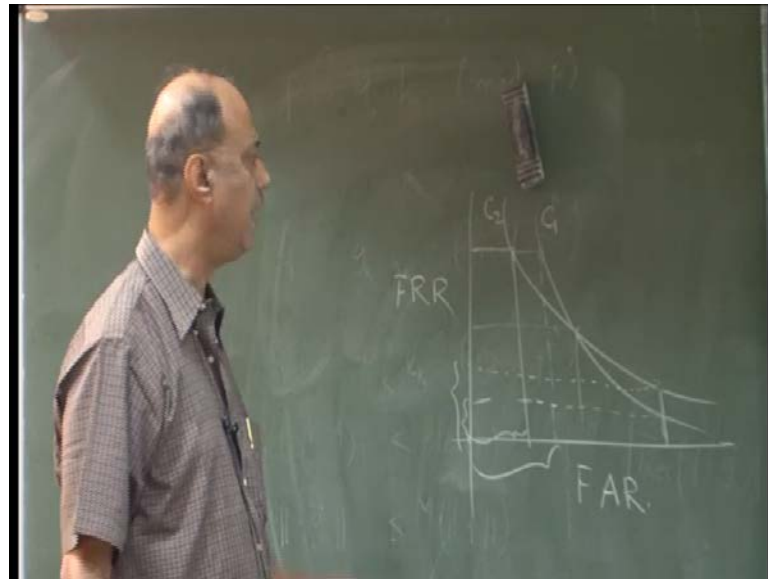
Now, sometimes what happens that you can tell that my system is better than the other system at this operating point; at this point if I consider that one is better than the other one, but for this diagram of course, it is not the case for this diagram that always you will find that my thing is worse than the matcher b. But you have to tell under what condition the thing is better than the other one and that case you have to tell that if I consider that m FR is this, FNMR is this, then my thing is different.

(Refer Slide Time: 05:52)



So, you have to give the two variables value that one is FMR, another one FNMR. If I consider that this is the case, then my thing is better than the other one. So, but as I told you that it is not **reality** really the two, the matcher you got ROC like that one is above the other one, this is not the case, you will find most of the case, the diagram will be like this.

(Refer Slide Time: 06:15)



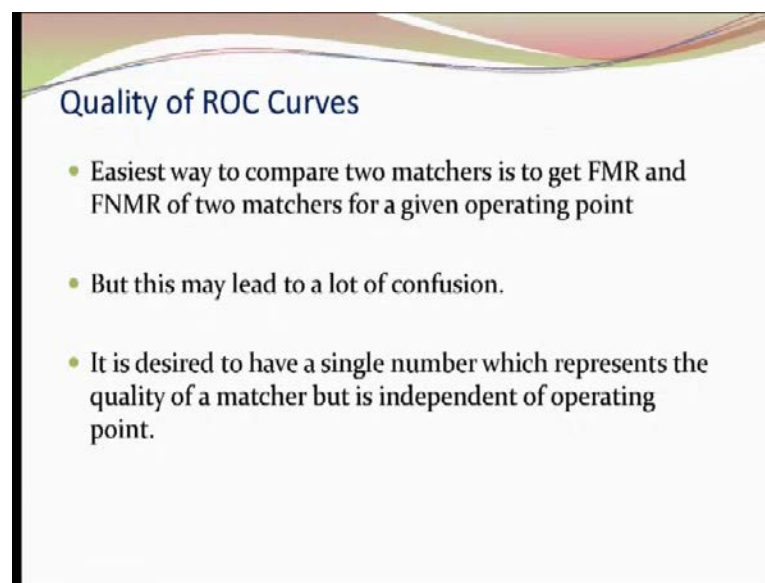
So, there will be intersecting point, and at this point, so both of them are having the equally **(())** the same FMR equals to FMR of a matcher and FMR of b matcher they are same. And in that case, this point you may consider as for comparison point for comparison purpose that ok, but that time you will not get the good result, because both of them are giving the same result. So, drawing conclusion based on the intersection point and will not be a correct method.

Now suppose this is curve one, and this is curve two; and I decide that my FMR, that is false acceptance rate has the higher priority than false rejection rate, false acceptance rate is because for some application, false acceptance rate is having the higher priority than false rejection rate. In that case, the C 2 will be better than C 1; C 2 will be better than C 1, because in that case, false acceptance rate is higher than the because I am keeping false rejection rate fixed, this is higher than this one, so you will be telling that C 2 is better, better matcher; am I right? Yes or no.

Now if I feel that no, I want the false rejection rate has the higher priority than the false acceptance rate than in this case, what happens under this this acceptance, false acceptance rate, I am fixing it, you will find that this is higher than this matcher C 2 is worse than matcher C 1, because this false rejection rate is higher. So, if it intersects then the you are in problem basically, you will be asking to the service, who wants the service, you will be asking which one you want, you want the false acceptance rate, then it should have the higher priority or false rejection rate should have the higher priority. Based on that you will be selecting that which curve is suitable for you, and you take the decision.

And if he tells that no, no, I do not care any one of them is ok or both of them are equally probable, then your problem is different, so that part will be discussed later on. But here at this stage what we are telling that false acceptance rate if you give the higher priority, then the C 2 is the better curve and if it give the false rejection rate has the higher priority, then C 1 is better curve.

(Refer Slide Time: 09:46)



The slide is titled "Quality of ROC Curves" and contains three bullet points. The background of the slide features a decorative wavy pattern in shades of green, yellow, and red at the top.

- Easiest way to compare two matchers is to get FMR and FNMR of two matchers for a given operating point
- But this may lead to a lot of confusion.
- It is desired to have a single number which represents the quality of a matcher but is independent of operating point.

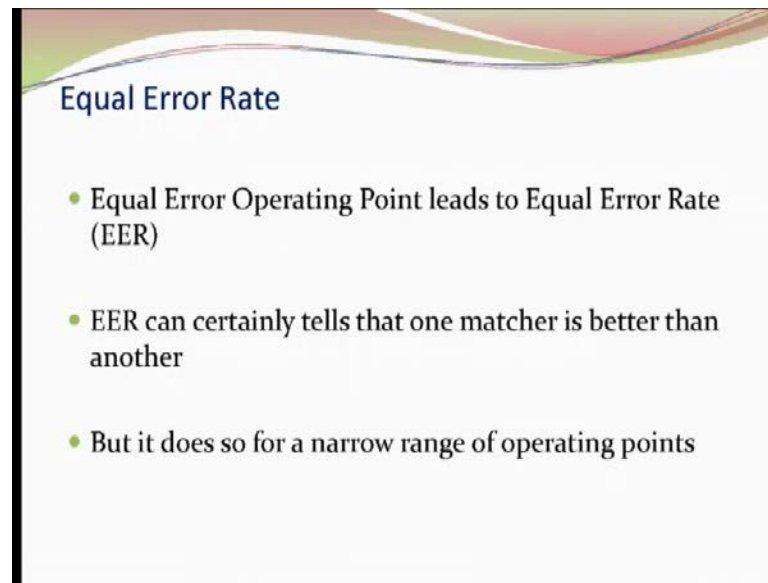
So, in order to tell that which one is better here? In that case, you have to tell the two parameters, one is earlier this, I was telling that no only one parameter is sufficient, the other will come automatically; if I tell F, a false acceptance rate, corresponding value of false rejection rate and you can obtain and corresponding value of T is also known to you.

But in this case, while you want to compare between the two methods not only you have to tell the false acceptance rate, you have to tell what is false rejection rate also, because the two matchers you have to tell corresponding value, so that you can justify which one even if you consider that my algorithm is good, who under this condition means under this false acceptance rate and false rejection rate, this is the best performance I am getting.

And you also will be telling that my algorithm will work very good under this false acceptance rate is this and false rejection, but how to compare this thing? This is a very difficult problem in that case, because you are getting only the two values FAR and FRR of each matchers, based on that you will be drawing the conclusion that is not possible, it is not you who will not understand, what is going on, because there two different parameters. You have to bring them under the same parameter which parameter if you assume that false acceptance rate. Let us make it fix and what is your false rejection rate and what is my false rejection rate, and then you compare.

But and then you possibly you will be getting the good results under this fixed false acceptance rate. Unfortunately, your matcher algorithm is not good for that false acceptance rate. It is good for some other false rejection acceptance rate. So depending upon, so this is not a good practice to go or it is not a good way to compare the two matchers under the keeping one of them fixed, because your algorithm may not good for that fix parameter and may be good for some other parameters.

(Refer Slide Time: 11:54)

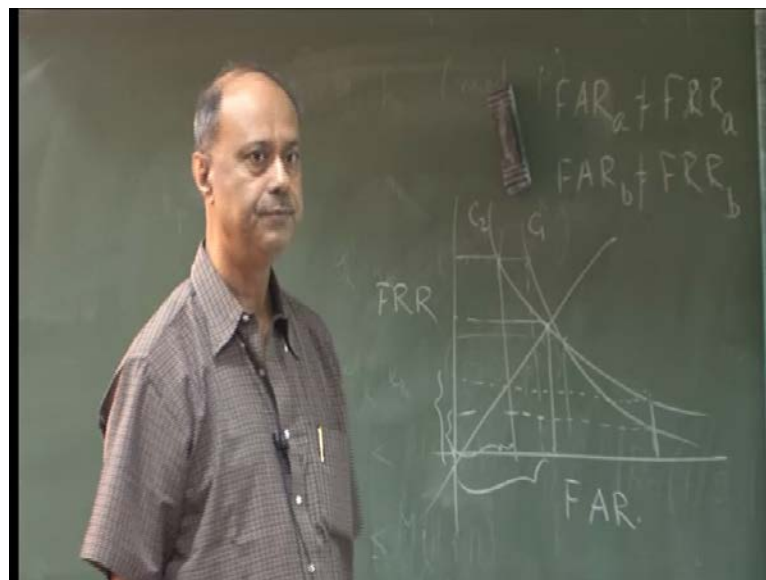


Equal Error Rate

- Equal Error Operating Point leads to Equal Error Rate (EER)
- EER can certainly tell that one matcher is better than another
- But it does so for a narrow range of operating points

So, that is the thing, what we want to tell here. So, what is the way out? Way out is that can I get that equal error rate? What does it mean? That equal error rate is nothing but under what condition, both false acceptance rate and false rejection rate is same.

(Refer Slide Time: 12:21)

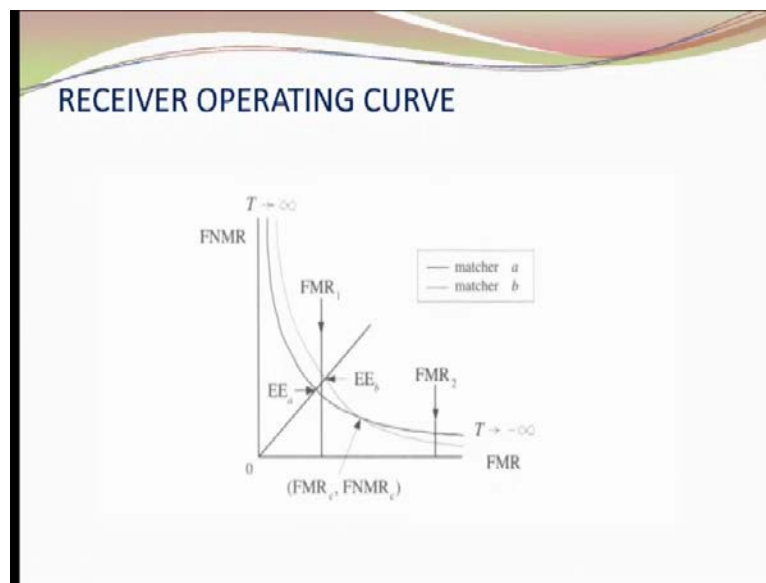


So, if I know one of them, FAR of a equals to F of a F of R of FRR of a; and similarly, FAR of b equals to FRR b, if I know this, then I ask, what is your false acceptance rate? And what is my false acceptance rate, because both of them is same, then whichever is smaller, that will be the better matcher, because now they say earlier degrees of freedom

was too many, because I have only one information false acceptance rate a and corresponding value of false acceptance rejection rate b . But a and b , these two are, this is not same by knowing one of them, by doing one of them or two of them, two of them, you are not able to draw any conclusion, which one is better? It is difficult.

So, what we are telling no. Let us make it fixed but, if I may fixed then you are facing another problem, because under that fixed FAR my algorithm may not be giving you the good FRR but, for some other FAR may be my algorithm is better. His algorithm is not better, because of this scenario so drawing conclusion is not possible or it is not correct what we are telling now. Let us the shown the condition where FAR and FRR are same, so my FAR and FRR is fixed same, your thing is also same. Now which one is smaller, that is the better algorithm. So, how to get that? This is nothing but line you are drawing at 45 degree under this condition, you will find that FAR and FRR are same. Now, smaller value of this is better than this. So, you will be telling that this algorithm is better than this algorithm.

(Refer Slide Time: 14:25)



So, you get equal error rate, and you will be drawing a line here, and this is equal error rate for these two. And you will be telling that if you use the equal error rate for your algorithm then matcher a is better than matcher b.

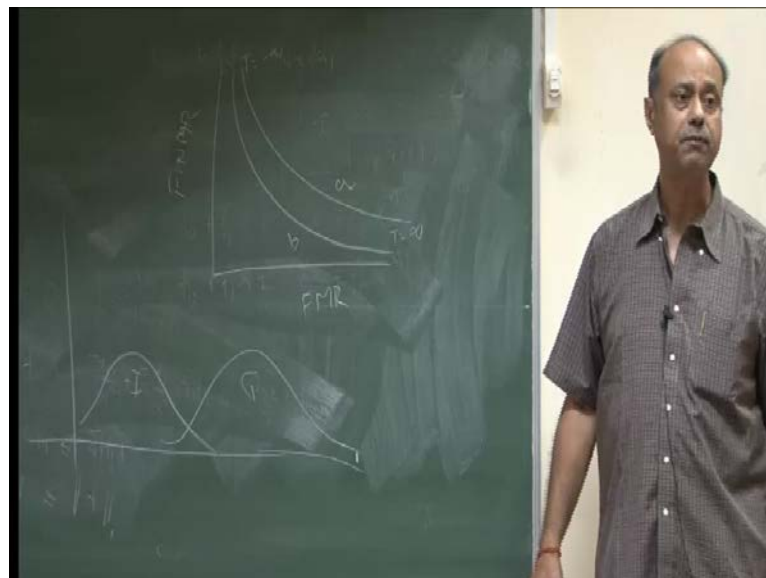
(Refer Slide Time: 14:45)

d-Prime

- Another way to judge the quality of a matcher is to measure how well the non-match score probability density $p_n(s)$ and match score probability density $p_m(s)$ are separated.
- A measure of this separation for a matcher is d' where
$$d' = (\mu_m - \mu_n) / \sqrt{(\sigma_m^2 + \sigma_n^2)}$$
where μ_m and σ_m^2 are mean and variance of the match scores of genuine users and μ_n and σ_n^2 are mean and variance of the non-match scores of mismatching users
- d' can be used to compare; but it has also some problem

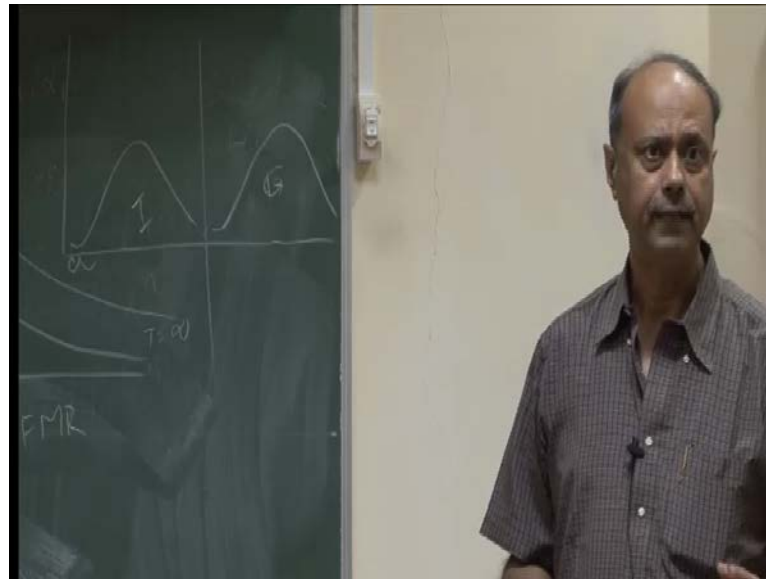
Now again if you see the single parameter FAR or EER is also not justifiable; again if you consider EER, there is also big problem, because you are drawing a conclusion, my matcher is not good or better than your matcher, because ER is different. Unfortunately ER is that, if you see the previous diagram ER is ER gives you this one and this one so very small difference is there. And this is such a small difference but for that small difference you are panelizing one matcher which is not correct.

(Refer Slide Time: 15:38)



So, what is your aim? Your aim should be that I have the two graph, one is genuine probability, another one imposter probability density function. Your aim, I will be telling that my algorithm should be the better one. If I can show that these two distributors are separated, widely separated, then this graph is here and this all imposter in the lower value, and all genuine who are far away from the imposter value.

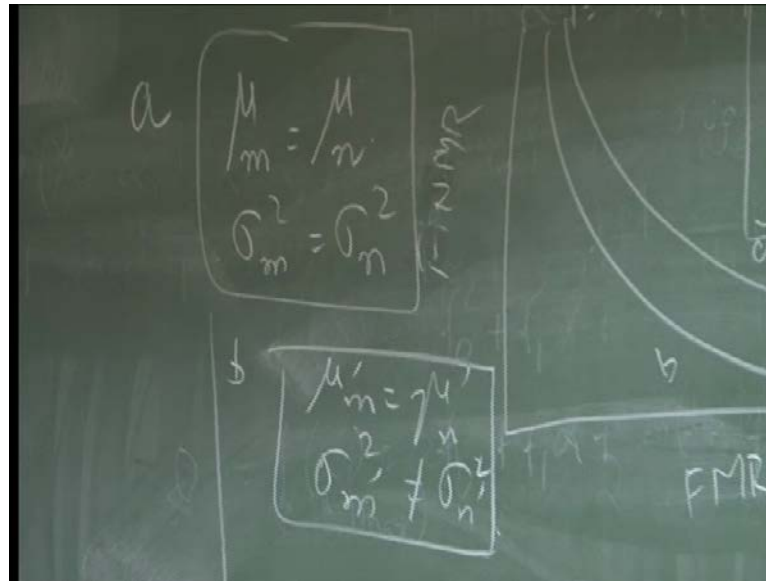
(Refer Slide Time: 16:15)



So, the best possible case will be imposter and this is genuine one, and I can have the threshold value like this. So, all of them will be rejected here and all of them will be accepted here, that is the best. So, how much they are separated? That separation value if parameter use a parameter that will be useful that is known as the d prime; d prime is nothing but, $\mu_m - \mu_n$, mean of these two distributions $\mu_m - \mu_n$ divided by $\sigma_m^2 + \sigma_n^2$. That will be variants of this distributions and variants of this distributions square root, if I divide it, then you can find out the d prime.

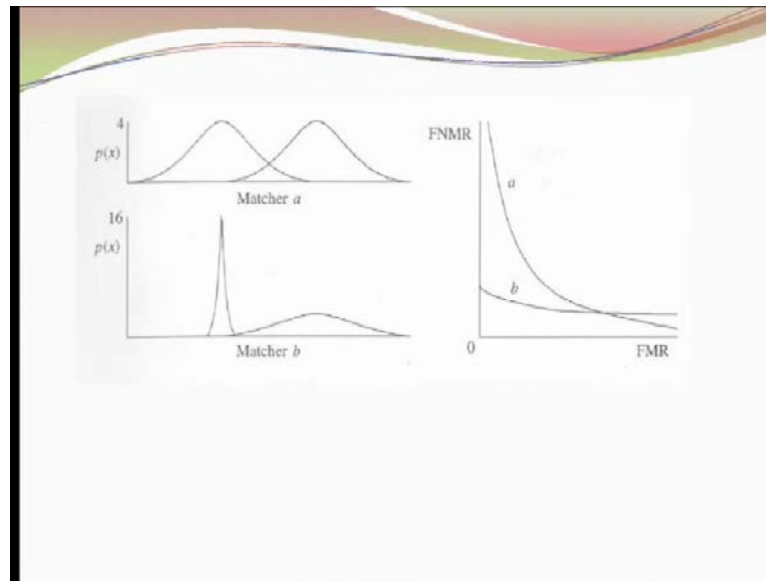
So, this d prime also can help you to determine, how much it is separated, how much it is so larger value of d prime, better is your matcher and you can tell that yes. This will separate it out from the imposter with this value of d prime; am I right? **Yes**.

(Refer Slide Time: 17:42)



So, here also you have some problem. Suppose, you have μ_m and μ_n , they are equal, and σ_m^2 equals to σ_n^2 ; this is for the matcher a; and in matcher b μ_m equal, μ'_m equals to μ'_n , and σ_m^2 equals not equals to σ_n^2 . So, both of them are having the same. This mean equals to this imposter mean, mean of genuine equals to mean of imposter mean, and here the variants of σ_m and variants of variants of m and variants of imposter, they are same and here but it is not the case in both the case d prime will be 0. What will be conclusion? You cannot draw any conclusion in that case, you will be telling both of them are alike; yes or no?

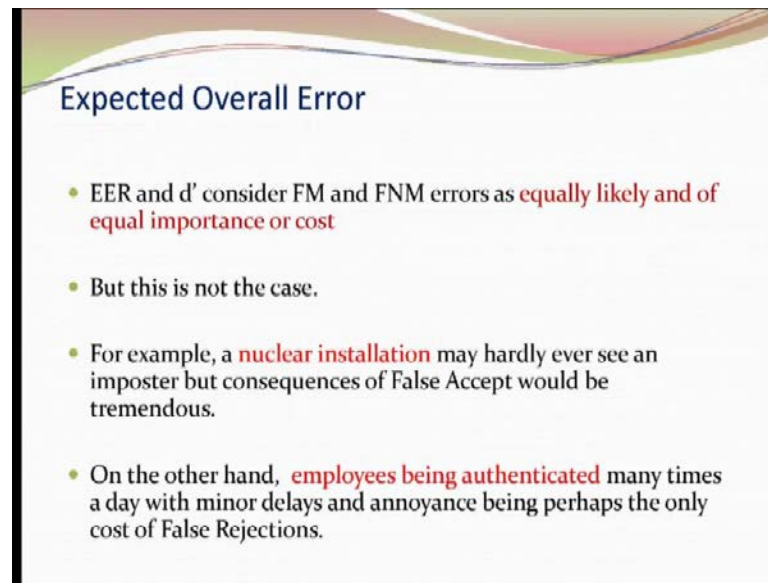
(Refer Slide Time: 18:39)



But if you see the graph here both of them are having the same mean; standard deviation is same, and your diagram will be look like this same mean, but standard deviation is different and its diagram may be like this false acceptance rate versus false rejection rate. So, you see that d' gives you 0, so you are telling that both of them are alike but in reality it is not in reality therefore, smaller value of FMR F a false acceptance rate the matcher b is giving good results.

So, anyway this is matcher b and matcher a. Under this condition, even if you are telling that d' is giving you the result, but you are not able to draw conclusion at this stage, but here you are finding that no matcher b is better than matcher a for small value of FMR, but and also for large value of FMR matcher a is better than matcher b.

(Refer Slide Time: 19:48)



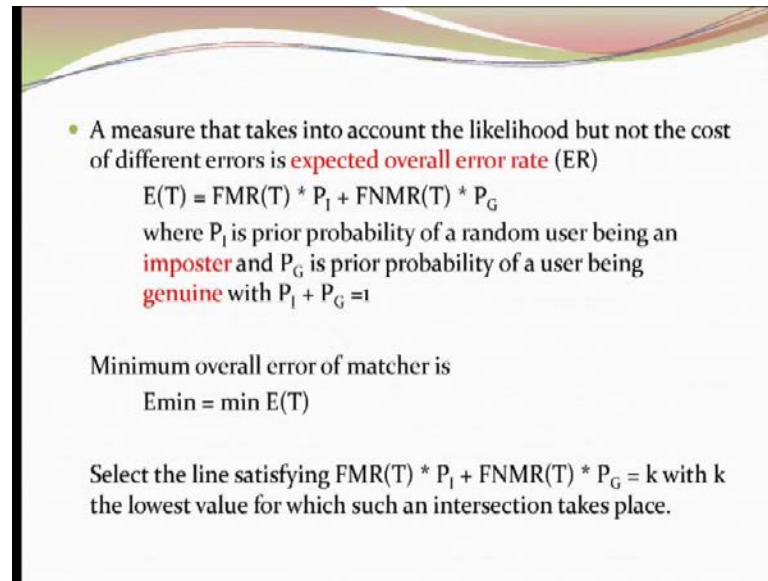
Expected Overall Error

- EER and d' consider FM and FNM errors as **equally likely and of equal importance or cost**
- But this is not the case.
- For example, a **nuclear installation** may hardly ever see an imposter but consequences of False Accept would be tremendous.
- On the other hand, **employees being authenticated** many times a day with minor delays and annoyance being perhaps the only cost of False Rejections.

So, you observe here one thing that both equal error rate and d' are giving you the equal importance to FAR and FRR. And also equal importance on the cost that what do you mean by cost? Cost means that making a false acceptance has some value, suppose in a nuclear plant I allow somebody falsely, it has tremendous effect, then it may create a havoc in the nuclear plant or if I stop somebody to come in entering genuine person, I am stop entering, genuine person in my lab. Several times that next time he will not come, he will tell that every time he is throwing me out, why shall I go there? Inconvenience will be there.

So, if he give the equal weightage or equal cost on this event that is also a big problem for us. So, you should not give equal the weightage to false acceptance and false rejection.

(Refer Slide Time: 20:59)



- A measure that takes into account the likelihood but not the cost of different errors is **expected overall error rate (ER)**
$$E(T) = FMR(T) * P_I + FNMR(T) * P_G$$
where P_I is prior probability of a random user being an **imposter** and P_G is prior probability of a user being **genuine** with $P_I + P_G = 1$

Minimum overall error of matcher is
 $E_{min} = \min E(T)$

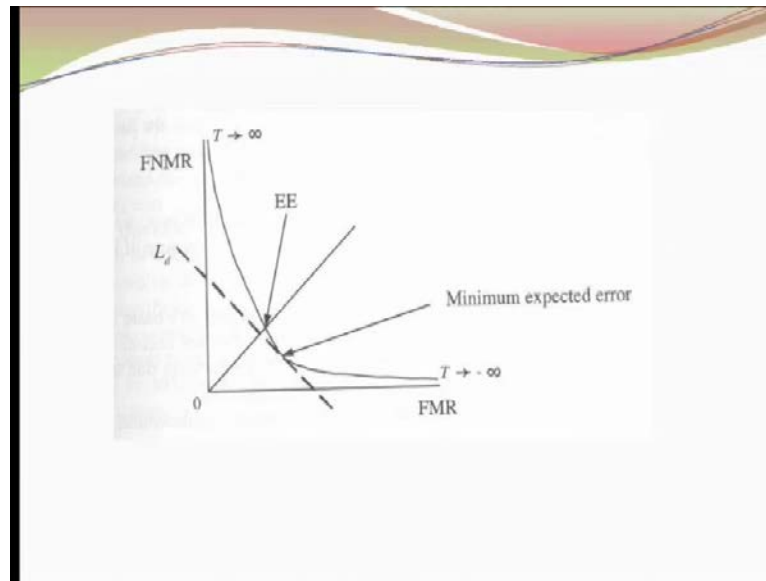
Select the line satisfying $FMR(T) * P_I + FNMR(T) * P_G = k$ with k the lowest value for which such an intersection takes place.

Now if it is the case, then we can define that expected overall error rate and the threshold point T is equal to false acceptance rate at T into probability of imposter plus false rejection rate at T into probability of genuine user. What it means? P_I imposter is the probability the random user is will be an imposter that probability that a person has come, he is random person and he is an imposter that I can find out.

And similarly, P_G is the probability that a random user is a genuine user. So P_I plus P_G must be equals to 1, so the false acceptance rate into the probability that a random user is an imposter plus false rejection rate at threshold T into probability that a random user is a genuine user that will give you the expected overall error rate, at the threshold value T .

Now your aim is to get such a T , such that this is minimum. So, you can draw now conclusion that if I know P_I and P_G , I know both of them and different threshold value I can find out what is the minimum value of $E(T)$; that is your minimum expected error rate. So, this is basically you will be getting $FMR(T)$ into P_I plus $FNMR(T)$ into P_G that will give you a value k , so the k is a line, this is a line.

(Refer Slide Time: 22:56)



So, you will be looking for that line, which touches that point; and this is your minimum expected error rate. So what I have considered? Now I have one thing you remember that I have considered P_I and P_G ; how to get the P_I and P_G ?

(Refer Slide Time: 23:21)

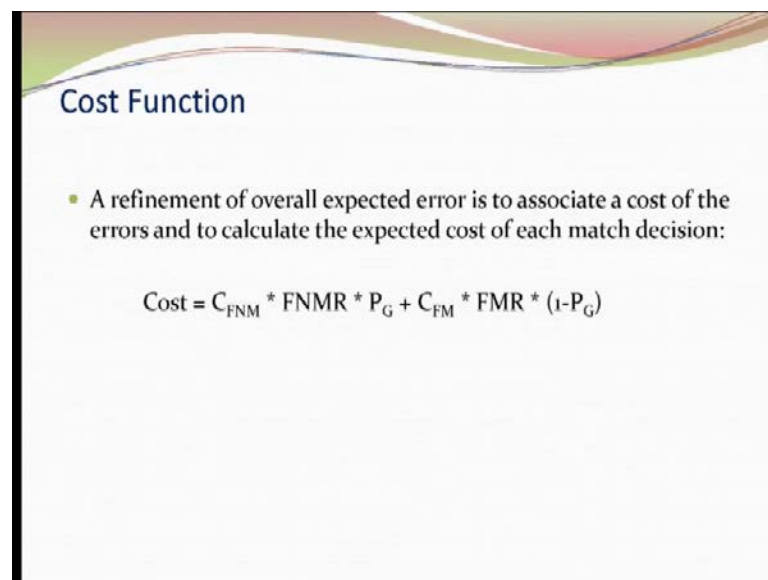
- It may be difficult to get the knowledge a priori about P_I and P_G
- A single measure of accuracy can be derived by setting $P_I = P_G = 0.5$
- But in general it is not the same as Equal Error Rate

That is a difficult problem, because in a lab, you have to ask somebody to come randomly, he will come and either he act as a genuine user or act as an imposter and if he has come as an imposter, then he should be matched with imposter that he has been considered as a false acceptance.

So, you have to find out that day distribution, you have to see; based on that, you will be obtaining your P I and P G. If I consider P I equals to P G equals to 0.5, because sum of this two must be equals to 1. That is nothing but your F a you are giving the weightage of half to both the side, which is nothing but half of false acceptance and half of false rejection you are adding to get the accuracy that means, sum of two errors divided by two gives you the accuracy.

And which is the normal practice; that you are giving equal importance to both of them but as I told you that it is not a correct approach, depending upon the application, you should decide, what is the weightage to be given on false acceptance and on false rejection rate?

(Refer Slide Time: 24:35)



Cost Function

- A refinement of overall expected error is to associate a cost of the errors and to calculate the expected cost of each match decision:

$$\text{Cost} = C_{\text{FNM}} * \text{FNMR} * P_G + C_{\text{FM}} * \text{FMR} * (1 - P_G)$$

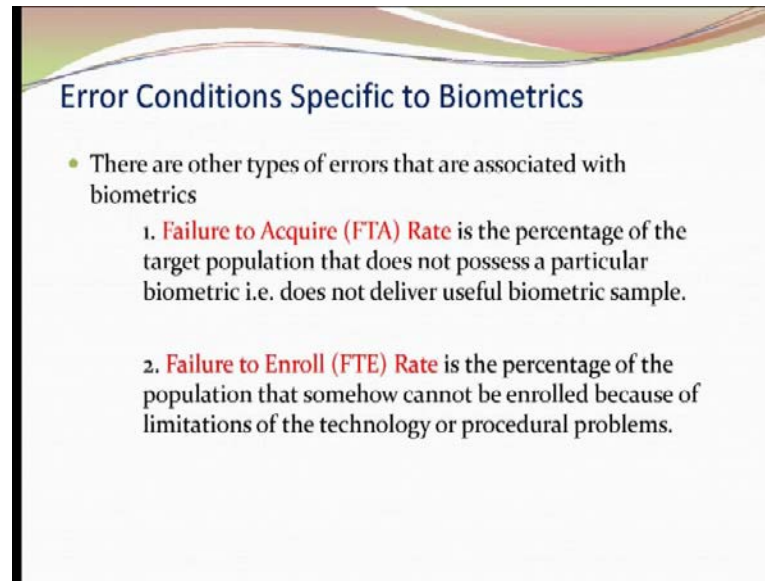
Generally, this should not be equals to equal error rate. Now you observe that we can improve further, if we introduce the term cost for false not match, and cost for false match. See every organization or every application, there is a cost for non match. If I stop you, well entering into my class room, and you are giving the data every time I am rejecting you, so you will be waiting outside; by that process, you will be losing ten minutes so that has a value that cost is there. So the cost you can obtain similarly, for this also false match also there is a cost, this cost you multiply, then you will be getting the cost, and you can determine such a curve, where it gives you the minimum cost.

So under what value of T for given P G P I P n a P C FM and C FNM and FAR and FRR you must be able to find a cost that T is important now for you; is it ok? Now, how to get FAR and FRR, because I told you, P G you are coming and I am testing randomly, and I am finding P G and P I that is possible. I know priory that cost of F m and cost of FNM this is also known to you, because I know that if somebody enters falsely, how much destroy, he will be doing in my organization. That is known or if somebody does not come or he is not allowed to enter the genuine person, then what is his wastage of time or cost?

So, this also can be estimated, but how to do these two? This is possible to get yes or no, because every organization has the value of assets, so that if somebody enters wrongly here to spoil I know what is the cost of my building or this area that is your cost. This is also known to you, how to get these two? Idea, is there any idea? You have designed a system and obviously while designing a system, you have tested it against whom? Tested it against a set of people or a set of subjects, which subjects what type of subjects? The subject should be similar to the application, suppose you are applying to students, your database will be tested and will be of students, because you should not use the database from the villagers. You know that environment where you are going to use it, that test data you have, and under this test data, what is FAR and FRR. You have already obtained that FAR and FRR will be used to determine your cost.

Now once you have deployed that machine, now you are getting the genuine data and now you there you monitor how much false acceptance, how how much false rejection rate at every instance of time, so correspondingly we change we update this one; so ultimately it will become a stable one. Yes or no? Is that clear? So, initially you have the lab environment based on that you do determine what is your FAR and FRR you deploy the system and keep monitor on the FAR and FRR part correspondingly. You upgrade your FAR and FRR, and then correspondingly your cost.

(Refer Slide Time: 28:22)



Error Conditions Specific to Biometrics

- There are other types of errors that are associated with biometrics
 1. **Failure to Acquire (FTA) Rate** is the percentage of the target population that does not possess a particular biometric i.e. does not deliver useful biometric sample.
 2. **Failure to Enroll (FTE) Rate** is the percentage of the population that somehow cannot be enrolled because of limitations of the technology or procedural problems.

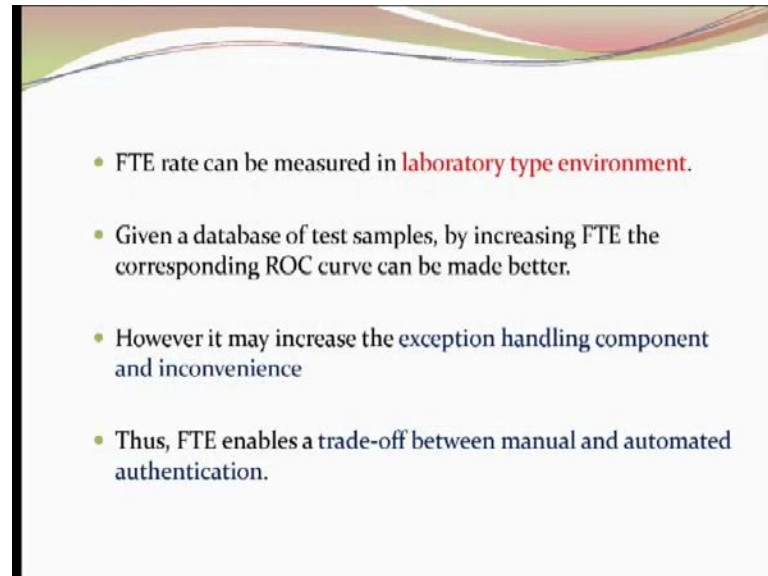
Now, whatever we are doing that FAR and FRR, but before that there are other issues, which generally we do not care, but if it is there they matters. There two type of errors still pending; one is failure to acquire; what it means? That you are a person and you do not have fingerprint or you do not have your eyes or you do not have finger. There is no guarantee that you have to have finger except the face, everything is variable. Only face should exist must exist, other part the finger may not be there.

Suppose I take the point finger, point finger may not be there or the fingerprints are missing that because of hard working, this has gone, nothing is there plain. So system will not be able to get any data from that. This is one type of error, and this is not the small thing; in **finger** case of fingerprint, it is coming 15 percent of the rural people. The number is very high. Failure to enroll, rate is also not a small number; this is even though, I have the fingerprint even though I have the iris, I stand in front of camera, but I am not able to give you the my iris data may be because of my bad day to provide the biometric data, and I am giving I have the fingerprint, I am giving the full pressure on the scanner but it is not giving me the result.

So these are the two types of errors failure exists in the **the** vendors, so what they do? They subtract this thing from the calculation, they immediately as soon as they find the quality image quality is poor and they delete it from their test data. Obviously their

accuracy will be very high, because they do not have the bad quality of data. All quality data is very good quality and result is very good.

(Refer Slide Time: 30:34)



So, what you can study the failure to enroll not the failure to acquire; failure to acquire you cannot do anything, because the fingerprint does not exist. So for that manual intervention or exception handling will be coming in between, but failure to enroll can be tested. What is the failure to enrollment rate? How can you obtain this? This can be done in the lab environment test.

How many you call people from different zones or different characteristics holders and tell them to provide the data, they provide the data based on that you can determine what is the failure to enrollment rate.

Now if you make very tight failure to enrollment rate, another any small problem you reject them. What is happening? Number of rejection and will be more, and that means manual intervention will be more, and what happens? If that gets the cost will be increased, because manually you will be enrolling them or you will be making them that **yes** or no. He is the identify the person or there is a chance of making mistakes will be more or chance of coming through in proper channel will be more, all those issues will be coming.

So, this failure to enrollment plays an important role right, to create the trade of relationship between the manual and automated system, if you make it very high failure to enrolment right, in that case what will happen that authentication will be very good, very good results you will be getting but manual intervention will be more, so cost will be more. And if it is a very small the value then it is very small, then that there is a chance of poor authentication technique and manual cost will be less.

So you have to decide what should be the appropriate value of FTE for in the case of iris. What you observe that 0.5 percent data is you know FT is very high, 0.5 percent; that means 0.005 that means what is a 0.5 means what? 5 out of 1000; 5 out of 1000 people, they are not able to give the data, it is a big number. So far 1000 people 5 people will be tested against exception through a exception handling.

(Refer Slide Time: 33:17)

Implications of Error Rates

- Why does it reject me?
 - Verification Protocol: Smartcard with Fingerprint
 - Suppose
 - Fingerprint system has 3% FRR,
 - 2000 people are requesting access per hour
 - One day means 16 hours
 - Number of people will fail to be verified = 960
- Why does it point to me?
 - Screening Protocol: Verification against Wish list
 - Suppose
 - Face system has 0.1% FPR,
 - Wish list database contains 25 faces
 - 500 people are requesting access
 - Number of people will be in the suspected list = 13
[25 * 500 * 0.001]

Now implication of error rates, that you have to understand carefully. Why does it reject me? That is one thing, that genuine person has come and say through verification mode. He has a smart card, he has given his biometric data and he has been rejected and he feels very bad, because I know there are because you know you go to the bank and you have signed on your cheque and they are telling no it is not your cheque or it is not your signature. You feel bad right, I it is me I have given my signature and you are telling it is not my signature.

So, how this happens you see here, fingerprints system has assume that three percent false rejection rate, three percent is very nominal, because it is much higher than... Let us assume that three percent false rejection, false rejection rate in a fingerprint system, and 2000 people are served per hour. It is a very small number, 2000 people are coming accessing the system per hour. There are 16 hours - 8 hours off and then the number of people will fail to verify this 3 percent into 4000 into 16 hours, so 916 is a roughly 1000 people will be rejected even though they are genuine. So, then the unfortunately you are one of them.

The next issue is coming why does it pointed to me, this is another problem in the you see the wish list area in the shopping mall you have decided; that there are 25 people whom I do not want to allowing in my shopping mall. And assume a system false, face system which has 0.01 percent false positive, which is also very nominal in a face database, face system. We see database contains 25 faces and 500 people entered into the shopping mall and out of this 500 people, because of one person 25 into 500 into 1 percent gives you 25 people. The 25 people will be accepted as a false accept. Unfortunately you are also one of them even though you are a nice person and your name should not be in the database, but it is showing that you are in that database.

(Refer Slide Time: 35:47)

Available Best Error Rates

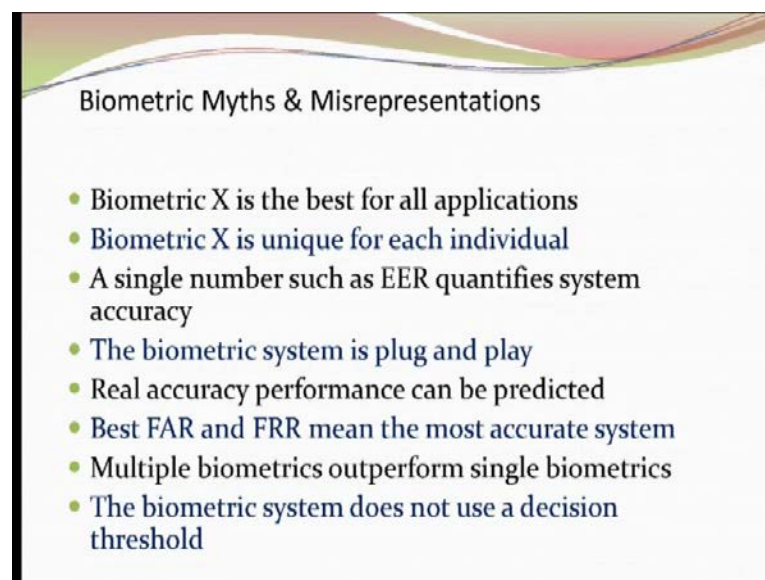
Trait	FRR (%)	FAR (%)
FINGERPRINT	3.0 - 7.0	0.001-0.01
FACE	10.0 - 20.0	0.1- 1.0
VOICE	10.0 - 20.0	2.0 - 5.0
IRIS	2.0 - 10.0	≥ 0.001
HAND	1.0 - 2.0	1.0 - 2.0
SIGNATURE	10.0 - 20.0	2.0 - 5.0

So, that is the error rate implications in traditional thing, if you find that generally fingerprint has the false rejection rate 3 to 7 percent after subtracting; all these (()) after

subtracting FTA and FTE false, then failure to a acquire and failure to enrollment after eliminating them. You will find that FRR is 3 to 7 percent and for that corresponding FAR is 0.0001, 0.01.

Face you see the very large value 10 to 20 percent, and with FAR 0.1 to 0.1 0.1 0.0 percent, voice is 10 to 20 percent with FAR is 2 percent to 5 percent, iris is 2 to 10 percent and it is more than 0.01 percent FAR. And hand is 1 to 2 percent 1 to 2 percent also there is signature is also very high, false acceptance rate corresponding to others. Where these are not standard things, these are not because we are also facing similar problem.

(Refer Slide Time: 36:53)



Now, biometric means and misrepresentation, this is one thing is important or we should discuss that biometrics X is the best for all applications. See, any biometrics one thing (()) cannot consider it is best for all. You have to understand carefully that what application it is what type of biometric is suitable, because you know lot many information is require, because cost is involve, accuracy is involve, and all then type of system you are going to consider; these are involve whether it is feasiuble apply, those system are also important.

Biometric X is the best for all applications, tt is not the correct thing; it is dependent upon the application you have to decide. Biometric X is unique for each individual, this is also not correct; remember one thing that we can tell for this set of people I am getting

the data where fingerprints are unique or face is unique. See I do not have the whole globe data I cannot draw the this type of conclusion first part.

Second part is that if it is unique, how come then error rate is three percent, four percent; that means there exist three percent people whose accuracy whose are matched with other persons, because of some other reasons. So, you cannot tell that this is perfectly unique, it is expected that this is unique. A single number such as ER quantifies system accuracy as I told you that even though we use ER or the cost or the minimum error rate, if expected error rate, but this will not give you the perfect feel about your system. You should tell more about it, the not only ER other behavior of the curve, you should explain clearly.

The biometric system is plug and play. It is not the case reason is that if it is a plug and play then lot many problem could have been solve, reason is first thing is that your system is dependent on the image or the how image quality how you are using them, and what part of this R O I you are extracting all those thing will be coming. Scanner may extract that X comma y size of image, another scanner is consider a comma b size of image, then how are you going to hand shake them, it is a (()) it is a plug and play is not the correct thing, it needs to lot of manipulation to make it a keep, real accuracy performance can be predicted.

It is also not possible, because your test data is based on the laboratory environment, and laboratory environment should the people gets enough time to give his data and also I can ask him to provide the data several time based on that you consider the best one, but in the real filed it is not the case. He gives the data and he is allowed to go, so this is not that accuracy, you can predict correctly; however that accuracy only we are taking into consideration for our calculations. Best FAR and FRR mean the most accurate system, that is also need not be because of the application you have to decided, what is the FAR and what is the best FRR?

Multiple biometrics outperform the single biometrics, that is the thing we want to in case, we are trying to do something that yes, multiple biometric if we have possibly error can be reduced. But, cost will be go up. This is another wrong way that decision threshold is not used in biometric system, because in the biometric system, since it is a

pattern recognition system you must use a threshold value, based on the threshold value you take the decision whether you are accepting or rejecting.

So, this is feature extraction can be used with any match engine. This is another wrong thing, because they are not mostly they are not interoperable. So, if I have one system, can I use your feature extraction algorithm for my matching algorithm, it will find that giving you the different problem. It will extract some features, but your matcher will not, because you do not know what is the size of the feature vector. What are the type of materials there in the feature vectors, you do not know, so as a result you are facing problem.

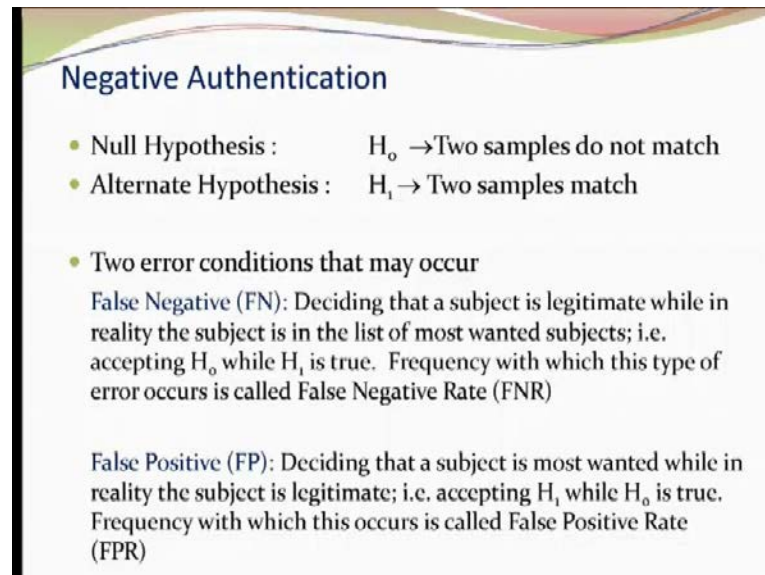
Large templates means better accuracy, need not be it depends upon the how accurate template. It is how accurate feature vectors, you have that is the important thing instead of the number of feature vectors - feature elements in the vector. face recognition prevents terrorism, it is also in a commonly we use face recognition for terrorism, that I see the face and then we draw the conclusion that **yes** he is the terrorist or not. But in reality it is not possible, because we cover the mass we use mass to cover our face and as a result this is not possible then your database may not contain the all the face data as a result the database is not a closed database. You will not be able to do this one.

Biometrics mean 100 percent security, but how it is possible? FAR and FRR are exist. They are not 0, if it is 0 then what you are telling is correct. This is not correct then it is a threat to privacy. This is a big problem, how can I give my biometrics data? How can you safe guard my biometrics data? Say, first thing is that you can encrypt, second thing is that your feature vectors, if I buy any means it has gone to somebody else Can he obtain a mask against your feature vectors, so inverse exist or not that is not known.

Biometrics sensors are unhygienic or otherwise harmful, tell me it is a unhygienic, do you expect that it is unhygienic or unhygienic ; this may be unhygienic, you do not know, because what happen that you once you go for the contact lens type thing that when you have to touch the lens. Then your predecessors who ever has come to give you data, you do not know what type of hand he has put. He can put the potassium cyanide on the top top of it and you have taken that, and you have given your fingerprint and you are also carrying the potassium cyanide with you. So, it is it may be harmful, it may be unhygienic. So, we should be that wherever the contact sensors you are using, you have

to be very careful. You have to justify that it is safe, otherwise it may create havoc in the society.

(Refer Slide Time: 44:01)



Negative Authentication

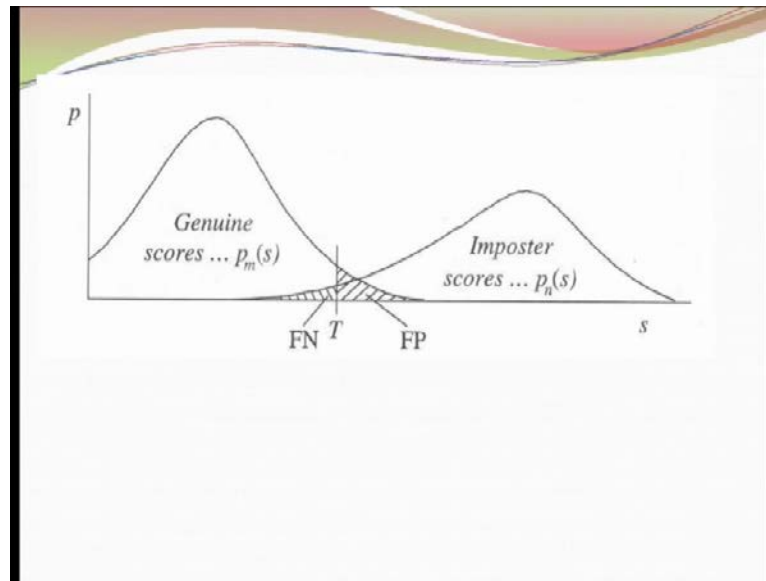
- Null Hypothesis : $H_0 \rightarrow$ Two samples do not match
- Alternate Hypothesis : $H_1 \rightarrow$ Two samples match
- Two error conditions that may occur
 - False Negative (FN): Deciding that a subject is legitimate while in reality the subject is in the list of most wanted subjects; i.e. accepting H_0 while H_1 is true. Frequency with which this type of error occurs is called False Negative Rate (FNR)
 - False Positive (FP): Deciding that a subject is most wanted while in reality the subject is legitimate; i.e. accepting H_1 while H_0 is true. Frequency with which this occurs is called False Positive Rate (FPR)

Now, whatever it is we discussed till now it was positive authentication. Now the negative authentication, you have the null hypothesis in the case of positive hypothesis. Positive authentication H_0 was two sample match, here it is the I do not want that two sample should match, because it is a wish list case.

So, null hypothesis is that two hypothesis do not match. Two samples do not match. Alternative one is that two sample match, so there are two errors in that case false negative means the two persons match, they suppose to be matched in reality. Because same person has come to enter the shopping mall, and it should have been matched with the database, but in practice he it was not matched. It was shown that no he is not the man with the database.

So, that is your false negative, and this frequency of this type of error is called false negative rate which is FNR false positive nothing but a subject which is not suppose to be matched with the database, but unfortunately he has been matched for false positive, and this for this frequency it is known as false positive rate.

(Refer Slide Time: 45:27)



So, this is the diagram. Imposter is this side and genuine is this side, pdf file and as we discussed in the previous case this is same as it is.

(Refer Slide Time: 45:40)

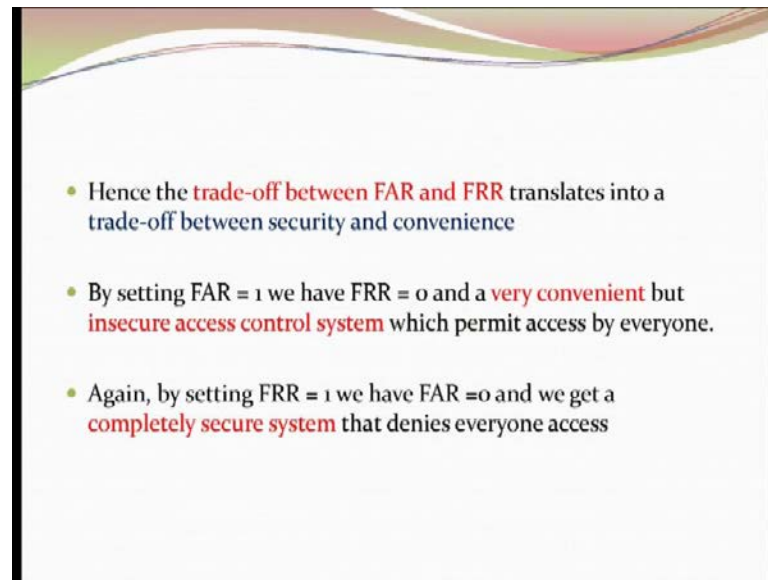
Trade-offs

- Two types of errors, that a verification system can make, have **different implications, and affect different class of people.**
- A False Accept into a secure system means an unauthorized person has gained access- a **breach of security**
- A False Reject means an authorized user has been denied access- which **does not impact security** but **inconveniences** the user and can have other implications by **preventing them from going about their business**

So, that two types of error; there verification system can make. One is FAR and FRR for F false positive and false negative and both of them has the different implications, and against different class of people. Now, false accept is a sequel system means an unauthorized person has been allowed, and that is mean a security beach false; accept in any sequel system means that unauthorized person is allowed what it indicates that is a

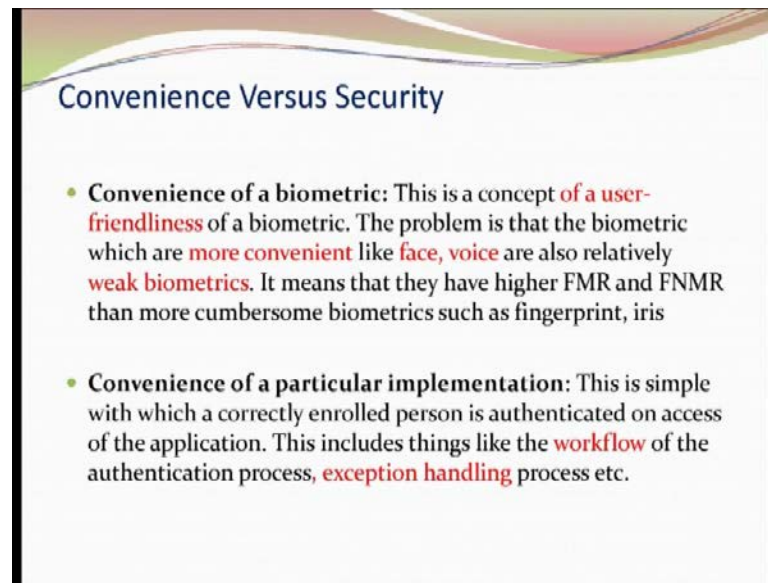
beach of security. While false reject is that, that an authorize person has been, he is a genuine person and you have stopped him to go there and as a result his inconvenience has been increased. So, here the implication is the beach of security and in this case inconvenience convenience problem.

(Refer Slide Time: 46:44)



So, that that trade of between FAR, FRR, FAR and FRR translate that trade-off between security and confidence. Now by setting false acceptance rate one, what it means? It means that false rejection is 0. That means everybody will be allowed to enter, that means what? Security chaos, and so it is a very insecure system. Everybody is allowed to enter insecure system. And FRR is one, false rejection rate is one; obviously false acceptance also will be 0, false rejection is one means I am stopping everybody and in that case confidence will be poor. Inconvenience will be very high. So FAR and FRR inversely related, these are two also inversely related. One is increasing, the other is decreasing.

(Refer Slide Time: 47:48)



Convenience Versus Security

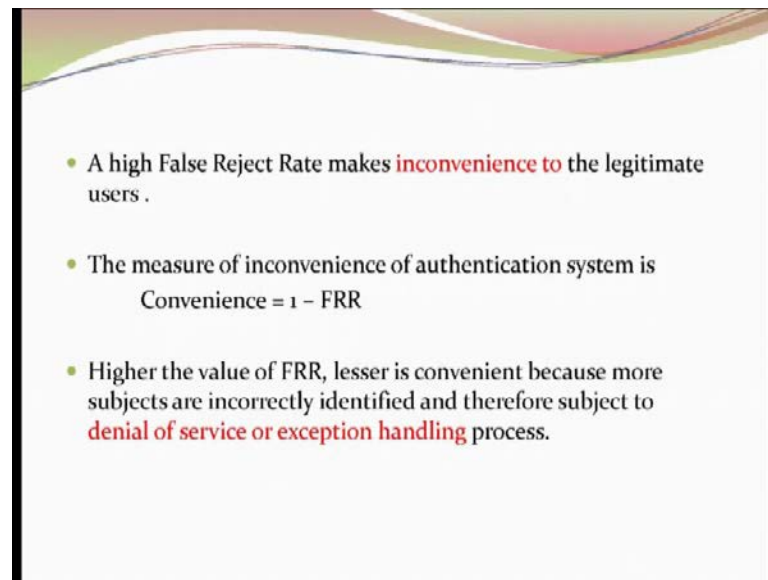
- **Convenience of a biometric:** This is a concept of a **user-friendliness** of a biometric. The problem is that the biometric which are **more convenient** like **face, voice** are also relatively **weak biometrics**. It means that they have higher FMR and FNMR than more cumbersome biometrics such as fingerprint, iris
- **Convenience of a particular implementation:** This is simple with which a correctly enrolled person is authenticated on access of the application. This includes things like the **workflow** of the authentication process, **exception handling** process etc.

Now, can I draw this statement convenience of the biometrics in nothing but the user friendliness. How much it is friendly with the user, how many time, how much I am allowing everybody to enter into the system easily.

So, one example is that face and voice, they are very easily available biometrics through which its very easy friendly environment through which anybody can enter. Obviously this will be more insecure system, but this is not the case with the fingerprint or iris. Where you are not allowing easily to enter into the system. Now, convenience of the particular implementation here it is not the biometric, it is the how are we are implementing the system.

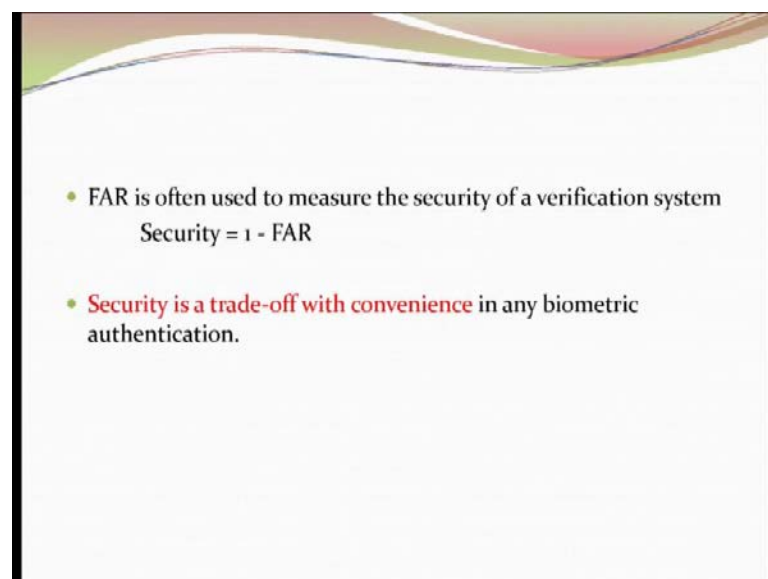
Now, in case of implementing the system, you have to see this part that how you are handling the exception. Because if you are making the secured system, lot many failure will be there or genuine people will be rejected. So, you have to do it manually and if you increase that false, if you are increasing the lenient one, your lenient means that user friendly one in that case what happens? That you are monitoring system should be accurate, this is so that you have allowed so many false people into the system. They should not create any problem.

(Refer Slide Time: 49:30)



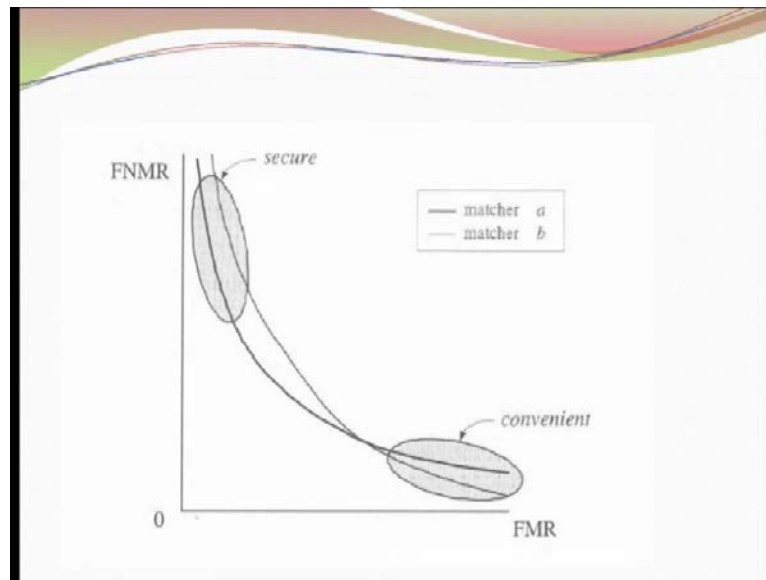
Both of them are important parameters, you have to understand carefully. So, as I know that high false rate makes inconvenience to the legitimate users. Convenience is nothing but 1 minus false rejection rate. You have false rejection rate is high convenience is 0, if false rejection is less convenience easy friendliness is high, so higher the value of FAR R lesser the convenience, because more subjects are incorrectly identified. And therefore, subjects or to denial or service or exception handling will be increasing.

(Refer Slide Time: 50:06)



And if false acceptance rate as I told that it is related with the security. If false acceptance is 0 then it is fully secured. If false security is false acceptance rate is zero, one then security is in problem. So, security and convenience are there is a trade or relationship between these two, because they are related with FAR and FRR.

(Refer Slide Time: 50:34)



So this is an example that you have the ROC curve. So, that I see the security secured system obviously this curve will give you the higher security, but if I see convenience one then this will give you the better result; that matcher a will be better than matcher b.

(Refer Slide Time: 50:57)

Cost Versus Security of Positive Authentication

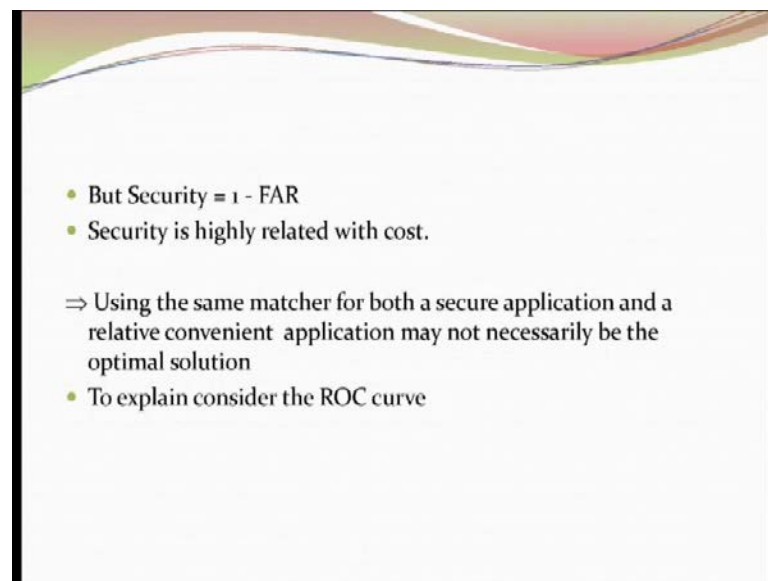
- There is a trade-off between **security and cost**
- By setting $FRR = 0$ and $FAR = 1$ we have a **very cheap but totally insecure system**
- By setting $FAR = 0$ and $FRR = 1$, the system is not accepting anybody and one has to rely **on costly human and manual labor**
- FRR can be used as some measure of cost of the system

$$\text{Cost} = FRR$$
- Higher the cost, more expensive an application because more subjects are incorrectly identified and therefore subject to denial of service or exception handling process

Now the question is coming the cost versus security. There exist a relationship trade of relationship between security and cost, because as security is high you have to pay more, because the manual intervention will be more, because security is high means false rejection rate will be there. And what will happen in that case you have to manual intervention will be there, you have to pay more. So, by setting false rejection rate is 0 false acceptance rate, we have to very cheap system very insecure system by setting FAR equals to. You are getting that anybody and who has to apply cost human manual (()). There lot many money you have to false acceptance, and you have to make it very tight. So too many people will be rejected, and manual intervention will be there, but cost will be involve there.

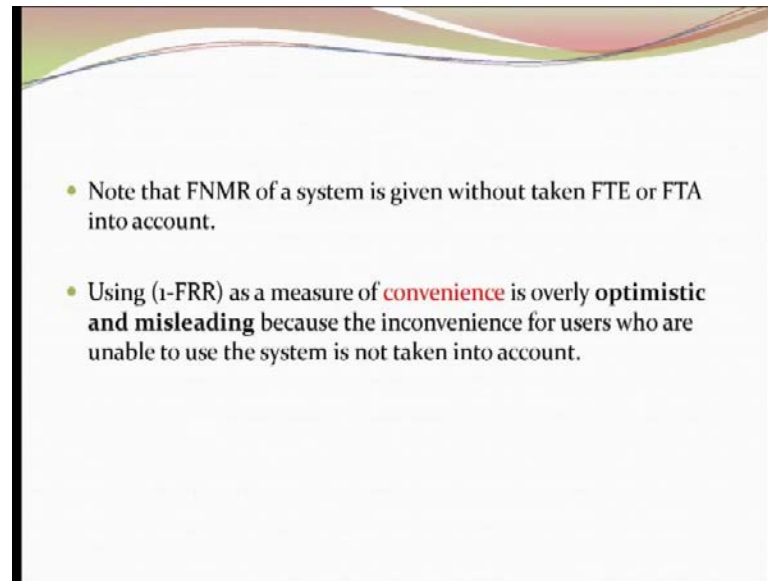
Cost is related with the false rejection rate, because as you are increasing the false rejection rate. They are genuine person and they will come for proving themselves that we are genuine person. So, they have to collect the data manually and verify them. Cost will be increased, so cost it is directly related to the false rejection rate.

(Refer Slide Time: 52:17)



But security is 1 minus FAR, as we have discussed; security is highly related with the cost, because as earlier, convenience was inversely related with the security, now we are telling that no security is related with cost. This I have already discussed; so this is the discussion on that.

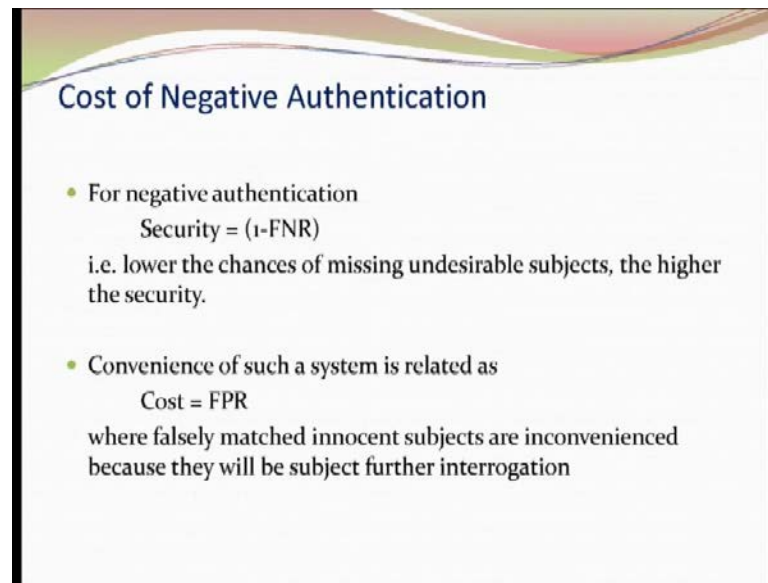
(Refer Slide Time: 52:42)



So you can assume that whatever we are obtain that we are not considering any FTA or FTE, so that is basically big problem. FTE and FTA and must be included in computing your security, because they are involve they are also see, I have no, I could not enroll them, because of their poor quality of image or because of some other reasons. But they are also in my database, and they will come with their information, and you have to allow them.

So, there exist some hidden cost, which we are not considering in computing, because computing our convenience or the security, I have use only FAR and FRR, but what about the case of FTE and FTA? So FTE and FTA, there value should be added to determine to the convenience factor or the cost factor, because they are also involved of that manual intervention, exception handling will be coming in between.

(Refer Slide Time: 53:54)

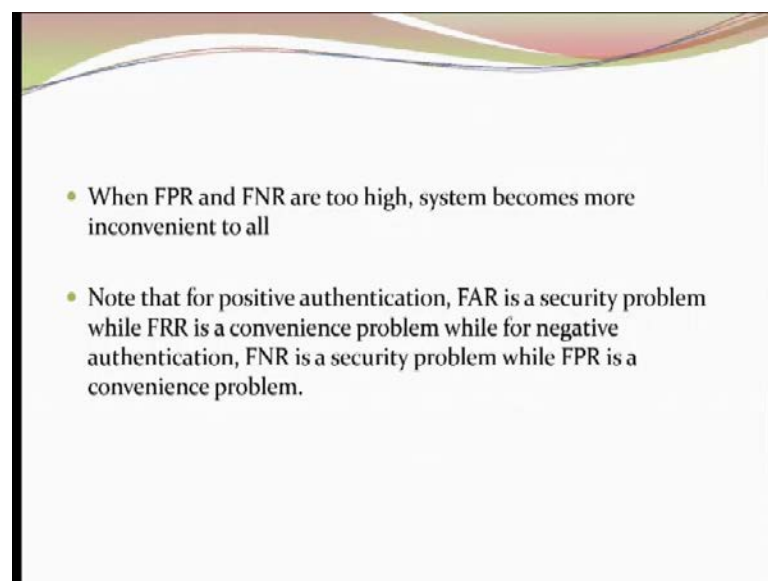


Cost of Negative Authentication

- For negative authentication
 $\text{Security} = (1 - \text{FNR})$
i.e. lower the chances of missing undesirable subjects, the higher the security.
- Convenience of such a system is related as
 $\text{Cost} = \text{FPR}$
where falsely matched innocent subjects are inconvenienced because they will be subject further interrogation

In the case of negative authentication, security will be just reverse of it is 1 minus FNR that lower the chance of missing undesirable subjects, the higher the security; and convenience is the cost of false positive. If false positive is there, then obviously that innocent people will be troubled, because they will be asked to stand up in a queue, and you interrogation will be there by police or by somebody else that how because they are the enlisted in the database. So false positive will keep increase the cost, and security is 1 minus FNR.

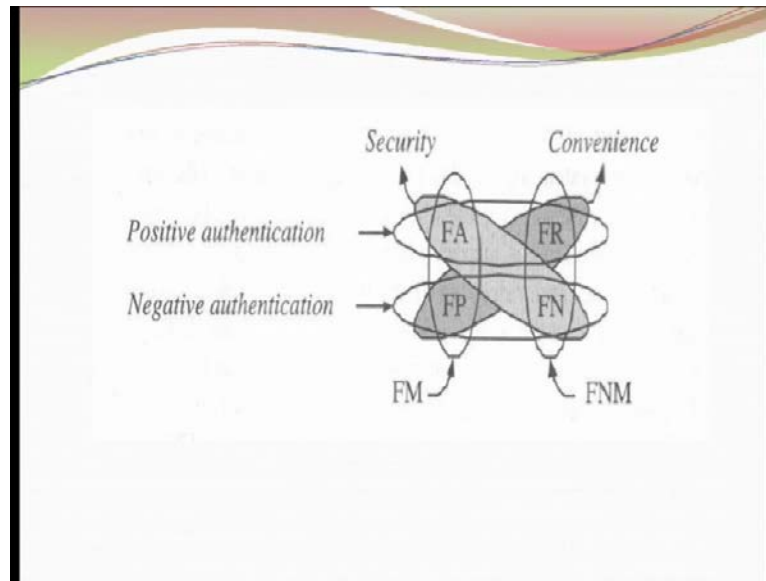
(Refer Slide Time: 54:36)



- When FPR and FNR are too high, system becomes more inconvenient to all
- Note that for positive authentication, FAR is a security problem while FRR is a convenience problem while for negative authentication, FNR is a security problem while FPR is a convenience problem.

So when a false positive and false negative rate is too high, system becomes inconvenient to all. And in the case of positive authentication, FAR is the security problem, while FRR is the convenience problem for the negative; while for negative authentication, FNR is the security problem and FPR is the convenience problem. If you see the diagram, I think that is still more.

(Refer Slide Time: 55:03)



What I am telling that this FA false acceptance and false negative, they are the security problem, and false rejection for the authentic[ation, positive authentication and false positive for the negative authentication, they are the convenience problem. So, this is your for security, and this is for your convenience; positive authentications security is FA, and convenience FR, for negative authentication false positive is the **security**, false positive is the convenience and false rejection is also convenience for the positive point. Similarly, later if I see the diagram that shall we can.

So, in the case of false match, it is false acceptance and false positive, they are giving you the false match; and and false not match is false rejection rate in the case of positive authentication, and false negative in the case of negative authentications. This diagram will explain everything, whatever I discussed till now.