Programming in C++ Prof. Partha Pratim Das Department of Computer Science and Engineering Indian Institute of Technology, Kharagpur

Lecture - 47 Type Casting and Cast Operators: Part I

Welcome to Module 32 of Programming in C++. In this module and the next couple of ones, we will discuss in depth about type casting and cast operators in C++.

(Refer Slide Time: 00:41)



So, the objective is to understand casting in C and C++.



This will be the overall outline. We have already made some small introduction to casting while we were talking about polymorphism, primarily the C kind of casting. And the fact that on a hierarchy up cast that is casting from a specialized class object to a generalized class type is safe; and downcast is not safe, it can lead to really serious errors we observed all of that.

So, we will take a quick recap on that. And then we will introduce what is known as cast operator in C++, there are four different operators we will discuss that. Ending with the dynamic casting scenarios, where casting is done based on the runtime type; and for that we will discuss another additional type id operator. So, for these modules, we will carry the same outline, thus parts to be covered in the specific module will be shown in blue as we have done here. And for that module the blue part will be visible on the left of your screen in every slide.

(Refer Slide Time: 02:16)



So, let us step back and ask for type casting, why should we cast. So, casting is primarily required to convert the type of an object type of an expression a function, function argument, a return value etcetera to another type. So, the basic concept of C++ is strongly rooted int type that is any variable, any constant, any object that we use has a well defined type or is expected to have a well-defined type. So, as we involve them in different expressions in different computations, there is a need that often for computing an expression, I need an object of a certain type, but at hand I have an object of a different type, and that is where I need to cast.

As we have seen in the context of C that a large number of type casting is done through implicit conversion or what can be said as silent conversion, these are standard C++ conversion; standard in the sense that conversions that are between known types and therefore, have already been enumerated by the language designer. So, there is a rule which dictates as to how I can implicitly convert an integer to a double or a double to an integer, whether I can convert a pointer to a certain type into a pointer to wide and so on.

And it will also have an implicit conversion, which is add on over C of user-defined conversions see you would recall that C did not exactly have user-defined type. So, the user-defined conversion was not a part of C, and that will come in here, and those can be

used as implicit conversions also.

The word implicit here particularly mean that when we do this kind of a conversion, we actually do not write anything extra in the source could to say that I am doing this kind of a conversion. But from the context of the expression, and the context of the object being used, you figure out that a conversion has happened. Naturally, in contrast, there is a whole lot of conversion, which is known as explicit conversion. Now of course, if we have the convenience of implicit conversion then why do we have explicit conversion? Explicit conversion is one where I explicitly write that I am making a conversion. So, if I x is of certain type T 1, I would try to write out that I would like to take x to another type T 2 convert it to another type T 2, and write it out explicitly in the source.

When I do that then I say I am doing explicit conversion and the reason I do explicit conversion could be many for example, it could be that implicit conversion is making a conversion which is not what I want. It is converting it to it into something which I am not comfortable with or it could be that implicit conversion finds that given an object or given an expression there are multiple possible conversions and it becomes ambiguous to which one should be taken.

And in those cases, and which are quite a few, you may need to make use of the explicit conversion. It is always usually, it is always a good practice to make conversions as explicit as possible, so that your intention in writing the code and what the compiler has understood for your intention becomes completely resolved, if you making an explicit conversion. Because you are writing it out in the part of the text if it is implicit then it depends on your understanding and the compilers understanding of that particular situation of conversion, and we will see how that may lead to different difficulties.

(Refer Slide Time: 06:50)



Now, to perform a conversion, to perform a type cast a compiler may now so far what the kind of things that you have seen, I have just try to illustrate one here that suppose we have written a function like this even in C. So, it is f is a function that takes two integers int i and int j and returns a double value and how does it return a double value by doing this. So, what all you have here is this is a nice illustration of the different conversions this is one where I make an explicit conversion. I say that i is an int, so i is of type int i want to take it to double.

Now what happens to j? j is also of type int. Now I do not say that actually I also need j to be taken to the type double, because I am doing a division where the first argument, the first operand is a double. And therefore, the second operand also needs to be double. So, there will be another conversion here which is implicit. So, this is a simple example where you have implicit as well as explicit conversion.

(Refer Slide Time: 08:21)



And as I have seen with a large number of my students, in students mind the thinking always is that i is an integer possibly its value was 2; j is another integer possibly its value was 3. And when I just convert it to double then I say that it becomes 2.0 or it will become 3.0 and so on and nothing actually happens, but that is not the case.



(Refer Slide Time: 08:40)

For the simple reason that thinks about a case of i being integer, it has a certain representation. So, this two has a representation which is an integer representation, which typically is a single field representation where you use some kind of a particular way to show the signed numbers. So, you use a particular style for doing that. And when you take it to double, then you get into a different representation which is a usually multipart, where you have a sign explicitly put in terms of whether it is positive or negative then you keep a certain characteristics of this and you keep a mantises part of it.

(Refer Slide Time: 09:36)



So, all that you say that if my number is 35.7, then I do not represent it; let me clear that (Refer Time: 09:36) if my number is 35.7, then I do not represent it as 35.7 or anything like that. I represent it in some normalize form, typically what the systems do these days that you convert it to 0.357 into 10 to the power 2, may be you will not do 10 to the power 2, you will possibly do this power in binary. I am just explaining it in decimal. So, the representation then becomes that my number is positive my fractional part is 0.357 that is it is a fraction which is less than 1 and then my power is 2. So, this representation gives me 35.7.

Now, certainly if I want to take it to integer, this will become 35, which is your 10s complement, but or 2s complement representation which is very different. Even if this

were 35.0, this was 0.350, this representation and the representation of 35 is very different. So, what is the consequence of that it is that you cannot just wish that through casting the type has changed, you actually have a different kind of value when you have changed it.

(Refer Slide Time: 11:01)



So, if I have i represented somewhere, and once I have cast it to double that same location cannot actually carry that values. So, I have a different location some location t 1, where I have the value of i equivalent value of i in the double representation. So, in many a times, we will see that compiler actually needs to allocate separate storage for this cast value and then initialize that with the casting process.

So, cast includes in many cases not always though, I will highlight where it will not need this it, but in many cases it needs to allocate memory where the temporary value, the cast value will be put, and it needs to perform computation to take the original value, and compute the value that have cast to. So, that is the basic idea of casting. And we will need to see how that happens well.



So, if I have a function like this then in all likelihood, the compiler actually is implemented a function something like this, this remains same. But it needs to do the conversion, so it defines a new variable temp i. And it is doing a conversion here, this is where the conversion is happening, this is like you can think of this as if like a constructor, so as if double as a constructor which is being initialized with an integer value i. So, temp i will become a new double variable having the value which is closest to the integer value i.

Similarly temp j, this will also need conversion will have a double corresponding to j and then you actually divide these two double numbers and give the result in double. So, this conversion, this conversion are what is required and these are the two fields which are required for the temporary storage. So, this is the scenario of casting that will happen in significantly in C, and that is what we keep on seeing when while we either use implicit casting, implicit conversion or when we use C style conversion in the process.

(Refer Slide Time: 13:12)



So, just to quickly recap we had seen a lot of C style conversion in module 26 before discussing getting into the dynamic binding. So, we have seen the usual you know integer double kind of conversion we have seen different other kinds of conversion also like we have seen that implicit conversion is not possible in certain cases like a pointed to double cannot be converted to int, but we explicit c style casting those things can be done.

(Refer Slide Time: 13:44)



We will see what are the consequences of that we saw different now the same rule if it is extended to C++ where classes come into the play, then we have seen that if there are two unrelated classes then mostly, no kind of implicit conversion between the objects are actually possible. But you can still use the C style of post casting to cast between the different pointers of these types.

(Refer Slide Time: 14:14)



We have seen the danger of doing this kind of casting; and because we have seen in this example earlier below that it is possible that I have a forced a casting, which is not to be cast, because it is a two unrelated classes. And the compiler allowed us to do that and then gave out certain garbage output.

 Module 32

 Ordue 32

(Refer Slide Time: 14:38)

So, going further on that, we saw the casting still may make some sense, if the classes are on a hierarchy. And in that, we saw that there could be up-cast or down-cast; up-cast is specialization to generalization; downcast is generalization to specialization. And we saw that up-cast is usually a relatively safe thing to do.

(Refer Slide Time: 15:02)



So, you can up-cast because when you are up-casting, you are using less information than what actually exist. So, we saw that these is up-cast are usually, OK; but if you down-cast then you are going from generalization to specialization, so you actually have less information and you are trying to infer more information and that gets risky.



(Refer Slide Time: 15:26)

So, coming to C++, the casting takes a completely different view significantly different view compared to C. In C, you had implicit cast and C style cast all of which know. And the two major factors is it often looses the type information in the several context, because you can force something into an expression into a certain type without really bothering about what happens to the representation, and the correctness of computation. And the casting really lacks the clarity of semantics. C++ has tried making this far more uniform and far more type consistent, so that your type always remains correct. So, in C++, you can perform fresh inference of types actually often without changing the value, because you often need to infer about type where you really do not want the value to change, you do not need the value to change, and we will show examples of that.

And in other cases, while you infer the type like you were doing in case of inferring an integer as a double; you need to change the value. And this can be done either through implicit computation, where you do not need to write how the change will happen. We have not written how integer is getting change to double or it could be situations where I will need to say the user will need to define how through the casting the change will happen. And with all that, in C++, the casting preserves type information in all context that is a very, very important factor which does not is not guaranteed in C. And it provides a clear semantics which is not also the case in C by making use of certain cast operators. So, there are fore cast operators, which you can explicitly write out to cast one value into another.

And the reason the cast operators are highly encouraged in comparison to implicit casting or explicit C style casting is a fact that if you have cast operators, you can see they have very peculiar kind of name const cast, static cast, reinterpret cast dynamic cast. So, if you have use them then you can easily do a grep kind of operation, grep is basically searching for a string in a file, those you have used Unix know this very well. So, you can use or you can use in your systems search operation, so if you just look for const cast, you will able to see exactly what are the points in your source code where you have used the constant casting mechanism. So, it has a lot of advantages.

And anything that you can cast by C style, you can cast more meaningfully using these operators. And therefore, once you have understood the casting in C++, you should not at

all use the C style casting at all in your code. If you need to use that, you will know that there is something wrong in your design or your understanding of coding.



(Refer Slide Time: 18:27)

Now, usually a casting is as I said that a variable is of certain type T 1 and I want to take it to some type T 2. So, the way it works is a cast operator takes. So, this is what we say is a source type and this is what we say is a target type. So, a cast operator takes the expression and specifies what is the target type, and that is how this conversion will happen, now it does not need to specify the source type because in C++ all expressions, all variables have a fixed type at any point they have a type.

So, knowing i, I will know the source type T 1 because i is of type T 1, but I need to know - what is a destination type, what is a target type and that will specify like this. And C++ has operators which can do casting at the static time or at the dynamic time.

(Refer Slide Time: 19:29)



So, therefore, cast operators which will discuss one after another. One is a first is a const cast operator; this operator is used to a basically deal with overwrite the const-ness or volatility that is the c-v qualification of an expression. So, it can convert take away the const-ness from a constant expression, it can add const-ness to a non-constant expression and so on. Static cast is done primarily through polymorphic is a polymorphic cast. So, if you are not on a polymorphic hierarchy then you can use static cast; and it often engages user defined casting, we will see that.

Reinterpret cast is a something between unrelated pointer type or between pointer and integer you can to reinterpret cast which is very, very risky; it is pretty much like the c style cast and should be very spearing used. And the most important of the cast is the dynamic cast, where you do the casting at the runtime based on the runtime time. So, we will see through all of these starting from the first one.



First is a const cast which is const cast converts between types of different c-v qualification c-v, you remember that is c stands for const v stands for volatile. So, if I have a variable declaration, I can qualify that by saying that it is either a constant that it will never change after it has been constructed or it could be volatile that is it can change at any time without our knowledge. So, whenever we have that we have the c-v qualification and const cast can change that is only const which can take away or remove the const-ness or volatility of an expression.

And usually does not perform any computation or change the value, because const-ness is more not in the value const-ness are in our understanding of the value. See if I have a variable i, and I know it is int now the variable at present the variable i at present may have value 5, now whether or not this value 5 can be changed in future is not a property of the value five. But it is understanding of or interpretation of this variable i as to whether it is const; if it is const then you are saying that this cannot be changed; if it is not const this can be changed.



And we have seen earlier also while discussing const-ness as a property that I may have a non-constant variable say i and I may have a constant reference to this variable. It is possible that I have a constant reference to this variable, which will mean that I cannot make assignments to this reference because this is a constant one while I can actually make changes to the variable itself. So, it can be looked at a multiple different ways.

(Refer Slide Time: 23:36)



Now, let us look at how does const cast work. So, I am taking a few simple situations for example, the first situation is having a print function. I mean do not worry about the functionality, it basically takes a char star pointer and prints that. And in the application, I have a char star pointer which is a const pointer because this is constant string; sample text is a constant string.

So, if I tried to do print c, you will get an error. Why will you get an error, because this c is a constant is a pointer to the constant data, you know the point constant is on this side, so the data cannot change, but the actual parameter the formal parameter in print is a non constant pointed to a non constant data, so it can change. So, this sees that if I pass c to this function then c might get changed; therefore, this call will not be allowed. So, you will get this is one kind of possible error that you can get.

(Refer Slide Time: 23:34)



Now if you want to actually really want to call this function, then you need to strip c of its const-ness. So, you can do it in this way; const cast is a name of the cast, c is the expression that you want to cast and this is your target type, your target type is char star. So, c was of type const char star, this was your source type. And now you are made it char star, so the type of this whole expression const cast char star c is char star is not const char star. So, it takes the const char star type stripes of the const and gives up a

char star. Now once it becomes char star then it is same as the char star type you have. So, now, you can call this function.



(Refer Slide Time: 24:35)

With this trick you can call the function in this particular case, and because you are using const cast anybody who reads to this code will immediately understand that you needed to strip the const-ness from c, and therefore, you have used it. Thing about another situations suppose you have a class a, and that has a const member function and a non-const member const member function cannot change the contents of the class, non-const member function can. So, if you have a constant object a then calling a dot get is fine because a dot get is a const member function it can be called with const or non const object, because it guarantees that it will not change.

Now think about calling a dot set a dot set this is a non constant member function it actually changes the object, and you are said that a is a constant object. So, a dot set is an error, because you are you cannot be allowed to change the value of a constant object which you can if you are using a non-constant member function. In terms of type basically this pointer of a, has a type const, it has a type which is of this. So, it says that this point is a constant pointer it points to a constant a object. And therefore, but set this function is a non-constant one. So, it needs a pointer of type A star const, but you have a

pointer of this type. So, you cannot pass it there, because if you can pass it there, then it can violate change anything and go ahead.

(Refer Slide Time: 26:34)



So, if you still need to call that function, what you can do we can strip the const-ness of the object from A and cast it, you do const cast and now you make it A ampersand that is a taking a reference you are creating a reference of A which is a constant object. And you are creating a non constant reference to that.

So, this resultant expression is an object which has this pointer which does not point to a constant object it points to a non-constant object now and therefore, you can it is to call the set function on this. But certainly you cannot do a similar thing using, you cannot strip the const-ness of the whole object altogether that is not allowed because that will actually mean that you need to construct a new object and do something else. So, this will still continued to be an error, but you can create a non-constant reference to the same object by doing the const cast.

(Refer Slide Time: 27:45)



So, this is the basic reason you need const cast for I have put in several other examples for you to practice. For example, this one show that how does it looks like if you do a C style. For example, here we had shown that you can strip const-ness of the string and make the function call go through.

Alternatively you can use C style and do this, I will strongly advice against doing this, because if you do this then anybody reading this will understand that you are striping const-ness of. If you write this then somebody reading this does not know why you are doing this C could have been you are doing this possibly because you know want to remove const-ness, but you are doing this may be because you has a void star pointer. So, you just want to make it understand it as a char star and so on. So, C style of casting does not give you any information and should be avoided.

Similarly, you could do this by creating a non-const reference to the constant object, and make this call we just saw that. You could also do this by a C style casting of this. You can cast the object to a non-const reference and then use that. I will again strongly advice against using this, because here it is clear that you are just removing the const-ness, here it is not clear as to what we are trying to do, it is not possible to make out, it will have to really understand that.

And the most dangerous thing is even if you tried to const cast, the whole object cannot be const cast. We cannot take a constant object and const cast to a non-constant object, because that will mean a completely different object. So, this even with const cast is an error, but surprisingly with c style this is permitted. So, you are actually doing something which is kind of illegal and goes against the basic premise of const-ness that you have built up. So, these should be strongly discouraged. So, please do not use this kind of things and all.

(Refer Slide Time: 29:45)



Finally, I have given some I say set of other examples here, showing you that if you have a constant member function then you can still make changes within that constant member function by changing the const-ness of this pointer within that member function. You can still do some tricks like that. Of course, do not try this with c style casting, because it will become very dangerous.

And just read through this line by line later on. I will just point out to one small piece of the code here. Here, I have defined a constant integer and initialize it with 3. I have defined integer pointer, but what I have done is have taken the address of this variable, which will be a pointer to a constant integer, because it is a constant integer. And therefore, I have striped the const-ness by the const cast.

So, I have j here which is constant and I have p j here which points which is nonconstant. So, this const-ness has allowed me to create this. And since this is nonconstant, I can assign through this because is pointing to a non-constant integer. So, I can do start p j and assign a value. If you do this after all this you accept undefined behavior you do not know what is going to happen. For example, all that we know that pointer points to a variable then if I print the value of the variable and if I deference and print the value from the pointer, I should get the same value.

So, if you look at this carefully if I print j and star p j, j is a variable p j is the pointer to this variable, star p j is certainly has to be same variable value. If I print that this is output that gets generated that this is 3 and this is 4. So, you are infer a quite a bit of surprise, if you use the const-ness to strip of the const cast to strip of const-ness arbitrarily. The reason this happens is a pretty simple that the compiler particular compiler that have used knowing that j is a const and three has actually replace 3 in this place. So, it does not j anymore, because it knows that it is const, so it cannot change and this one is a changing value.

(Refer Slide Time: 32:24)



So, when it had to do this conversion it has silently created another location, so p j actually. So, j is here which has three when this conversion was done p j actually does

not point here, but it points to a new temporary location whose value is initialized with the value of three and then the changes happened within that. You can clearly see that they give you different results. So, go through this, you will get further details, you can learn that function pointers const cast cannot be removed from the function pointers and so on.

(Refer Slide Time: 33:00)



To summarize, we have tried to understand the basic process of casting in C and C++. And particularly explained cast operators, a basic structure of the cast operator and discuss the evils of C-style casting in a number of examples. And particularly, we have taken a look into the const cast operator.

In the next module, we will take up the other cast operators and move on.