Workshop on Essential Abstractions in GCC

Introduction and Opening Remarks

GCC Resource Center (www.cse.iitb.ac.in/grc)

Department of Computer Science and Engineering, Indian Institute of Technology, Bombay



July 2009

Introduction: Outline

- Genesis and Objectives of GCC Resource Center
- Motivation behind the Workshop
- Philosophy and Pedogogy of the Workshop

1/29

July 09

Part 1

Genesis and Objectives of GRC

• An Informal Group

2/29

A Desire

July 09



2/29

► CSE faculty members at IITB: Uday Khedker Amitabha Sanyal

Supratim Biswas

A Desire

How Did It All Begin?

- An Informal Group
 - CSE faculty members at IITB: Uday Khedker
 Amitabha Sanyal
 Supratim Biswas
 - Reasonably long and deep experience of research in compilers
- A Desire

July 09



How Did It All Begin?

- An Informal Group
 - CSE faculty members at IITB: Uday Khedker
 Amitabha Sanyal
 Supratim Biswas
 - ► Reasonably long and deep experience of research in compilers
- A Desire

Performing research grounded in theory and corroborated by empirical evidence

How Did It All Begin?

- An Informal Group
 - CSE faculty members at IITB: Uday Khedker
 Amitabha Sanyal
 Supratim Biswas
 - ▶ Reasonably long and deep experience of research in compilers

A Desire

Performing research grounded in theory and corroborated by empirical evidence

- ► Exploring research issues in real compilers
- Demonstrating the relevance and effectiveness of our research in real compilers

A Modest Start in 2003...

Introduction: Genesis and Objectives of GRC

3/29

• Our Guinea Pigs

Our Tool of Experiment

Essential Abstrations in GCC

GCC Resource Center, IIT Bombay

3/29

Our Guinea Pigs

Essential Abstrations in GCC

GCC GCC Resource Center, IIT Bombay

Introduction: Genesis and Objectives of GRC

A Modest Start in 2003...

• Our Tool of Experiment The Gnu Compiler Collection

A Wodest Start III 2003...

Introduction: Genesis and Objectives of GRC

- Our Tool of Experiment The Gnu Compiler Collection
 - Compiler generation framework
 Stable compiler generated for several dozen targets

Our Guinea Pigs

July 09

- Our Tool of Experiment The Gnu Compiler Collection
 - Compiler generation framework
 Stable compiler generated for several dozen targets
 - Millions of users
- Our Guinea Pigs

A Modest Start in 2003...

- Our Tool of Experiment The Gnu Compiler Collection
 - Compiler generation framework Stable compiler generated for several dozen targets
 - Millions of users
- Our Guinea Pigs

Several unsuspecting M.Tech. students, external B.E. students, and project engineers

And Then in 2007...







http://www.cse.iitb.ac.in/~uday/gcc-workshop/?file=intro

Organised by

Centre for Formal Design and Verification of Software and Dept. of Computer Science & Engg., IIT Bombay.

Home

This workshop is a 3-day instructional

workshop (and not a forum for contributed presentations) and involves lectures and laboratory exercises aimed at providing details of the internals of GCC which is an acronym for GNU Compiler Collection. It is the de-facto standard compiler generation framework on GNU/Linux and many other variants of Unix/Linux on a wide variety of machines. In the last 20 years of its existence, it has

acceptability. The focus of this workshop is different

seen a rapid growth and wide

News

(15 Aug 07). Our paper on incremental

construction of machine descriptions has been accepted for presentation at the GREPS 2007 workshop. This paper describes the methodology which was taught in our workshop at IIT Bombay.

→ Google

4/29

S

is available at the downloads page. (10 July 07). There have been some delays in organizing the slides page and in sending certificates. I am tied up with a couple of

deadlines. I hope to do the needful soon enough, perhaps on this week end. Will send a mail to all

(23 July 07). Slides and other workshop material

participants once this is done. Prof. Uday Khedkar Sun Jan 25, 8:33 PM

GCC Resource Center, IIT Bombay

File Edit View History Bookmarks Tools Help

Home

GCC

About GCC

workshop

workshop

Schedule

Participation

Important Dates

Other workshops on

Take-aways from this

Reaching IIT Bombay

Applications Places System

The focus of this

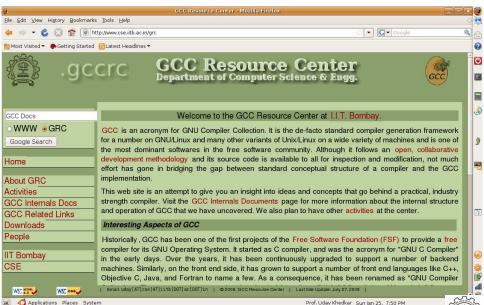
Thanks to small seed grants from IITB and IBM Faculty Award...

Introduction: Genesis and Objectives of GRC

And then in 2008...



And then in 2008...



Introduction: Genesis and Objectives of GRC

 A generous grant from the Department of Information Technology, Ministry of Communication and Information Technology, Government of India.

6/29

July 09

July 09 Introduction: Genesis and Objectives of GRC 7/29

Plan for July 2009...



GCC Resource Center, IIT Bombay

8/29

Objectives of GCC Resource Center

1. To support the open source movement Providing training and technical know-how of the GCC framework

to academia and industry.

Objectives of GCC Resource Center

- To support the open source movement
 Providing training and technical know-how of the GCC framework to academia and industry.
- To include better technologies in GCC
 Whole program optimization, Optimizer generation, Tree tiling based instruction selection.

Objectives of GCC Resource Center

- 1. To support the open source movement Providing training and technical know-how of the GCC framework to academia and industry.
- Whole program optimization, Optimizer generation, Tree tiling based instruction selection.

2. To include better technologies in GCC

- 3. To facilitate easier and better quality deployments/enhancements of GCC
 - Restructuring GCC and devising methodologies for systematic construction of machine descriptions in GCC.

Objectives of GCC Resource Center

- To support the open source movement
 Providing training and technical know-how of the GCC framework to academia and industry.
- Whole program optimization, Optimizer generation, Tree tiling based instruction selection.

2. To include better technologies in GCC

- To facilitate easier and better quality deployments/enhancements of GCC
 Restructuring GCC and devising methodologies for systematic construction of machine descriptions in GCC.
- To bridge the gap between academic research and practical implementation
 Designing suitable abstractions of GCC architecture

Broad Research Goals of GCC Resource Center

- Using GCC as a means
 - Adding new optimizations to GCC
 - Adding flow and context sensitive analyses to GCC (In particular, pointer analysis)
- Using GCC as an end in itself
 - Changing the retargetability mechanism of GCC
 - Cleaning up the machine descriptions of GCC
 - Systematic construction of machine descriptions
 - Facilitating optimizer generation in GCC

GRC Training Programs

Title	Target	Objectives	Mode	Duration
Workshop on Essential Abstractions in GCC	People interested in deploying or enhancing GCC	Explaining the essential abstractions in GCC to ensure a quick ramp up into GCC Internals	Lectures, demonstrations, and practicals (experiements and assignments)	Three days
Tutorial on Essential Abstractions in GCC	People interested in knowing about issues in deploying or enhancing GCC	Explaining the essential abstractions in GCC to ensure a quick ramp up into GCC Internals	Lectures and demonstrations	One day
Workshop on Compiler Construction with Introduction to GCC	College teachers	Explaining the theory and practice of compiler contruction and illustrating them with the help of GCC	Lectures, demonstrations, and practicals (experiements and assignments)	Seven days
Tutorial on Demystifying GCC Compilation	Students	Explaining the translation sequence of GCC through gray box probing (i.e. by examining the dumps produced by GCC)	Lectures and demonstrations	Half day

GRC Training Programs

Title	Target	Objectives	Mode	Duration
Workshop on Essential Abstractions in GCC	enhand	4, and 5 July, 2009 Γ Bombay, Mumbai	Lectures, demonstrations, d practicals (experiements and assignments)	Three days
Tutorial on Essential Abstractions in GCC	People interested in knowing about issues in deploying or enhancing GCC	Explaining the essential abstractions in GCC to ensure a quick ramp up into GCC Internals	Lectures and demonstrations	One day
Workshop on Compiler Construction with Introduction to GCC	College teachers	Explaining the theory and practice of compiler contruction and illustrating them with the help of GCC	Lectures, demonstrations, and practicals (experiements and assignments)	Seven days
Tutorial on Demystifying GCC Compilation	Students	Explaining the translation sequence of GCC through gray box probing (i.e. by examining the dumps produced by GCC)	Lectures and demonstrations	Half day

Part 2

Motivation Behind this Workshop

• To understand GCC well :-)

Introduction: Motivation Behind this Workshop

Motivation Behind this Workshop

Introduction: Motivation Behind this Workshop

Motivation Behind this Workshop

11/29

• To understand GCC well :-)

Reasonably quickly

Essential Abstrations in GCC

GCC Resource Center, IIT Bombay

Why is Understanding GCC Difficult?

Some of the obvious reasons:

Open development model

- Comprehensiveness
 GCC is a production quality framework in terms of completeness
 and practical usefulness.
- Leads to heterogeneity of the design.

 However now the main plan is vetted by the steering committee.
- Rapid versioning
 GCC maintenance is a race against time.

C, C++, Objective-C, Objective-C++, Java, Fortran, and Ada
Processors supported in standard releases:

Introduction: Motivation Behind this Workshop

Comprehensiveness of GCC 4.3.1: Wide Applicability

13/29

July 09

Comprehensiveness of GCC 4.3.1: Wide Applicability

Introduction: Motivation Behind this Workshop

- Input languages supported: C, C++, Objective-C, Objective-C++, Java, Fortran, and Ada
- Processors supported in standard releases:
- Common processors:
 - Alpha, ARM, Atmel AVR, Blackfin, HC12, H8/300, IA-32 (x86), x86-64, IA-64, Motorola 68000, MIPS, PA-RISC, PDP-11, PowerPC, R8C/M16C/M32C, SPU, System/390/zSeries, SuperH, SPARC, VAX

Comprehensiveness of GCC 4.3.1: Wide Applicability

Introduction: Motivation Behind this Workshop

- Input languages supported:
- C, C++, Objective-C, Objective-C++, Java, Fortran, and Ada
- Processors supported in standard releases:
 - Common processors:
 - Alpha, ARM, Atmel AVR, Blackfin, HC12, H8/300, IA-32 (x86), x86-64, IA-64, Motorola 68000, MIPS, PA-RISC, PDP-11, PowerPC, R8C/M16C/M32C, SPU, System/390/zSeries, SuperH, SPARC, VAX
 - Lesser-known target processors:
 - A29K, ARC, ETRAX CRIS, D30V, DSP16xx, FR-30, FR-V, Intel i960, IP2000, M32R, 68HC11, MCORE, MMIX, MN10200, MN10300, Motorola 88000, NS32K, ROMP, Stormy16, V850, Xtensa, AVR32

Comprehensiveness of GCC 4.3.1: Wide Applicability

Introduction: Motivation Behind this Workshop

- Input languages supported:
 - C, C++, Objective-C, Objective-C++, Java, Fortran, and Ada
- Processors supported in standard releases:
 - Common processors:
 - Alpha, ARM, Atmel AVR, Blackfin, HC12, H8/300, IA-32 (x86), x86-64, IA-64, Motorola 68000, MIPS, PA-RISC, PDP-11, PowerPC, R8C/M16C/M32C, SPU, System/390/zSeries, SuperH, SPARC, VAX
 - Lesser-known target processors: A29K, ARC, ETRAX CRIS, D30V, DSP16xx, FR-30, FR-V, Intel
 - i960, IP2000, M32R, 68HC11, MCORE, MMIX, MN10200, MN10300, Motorola 88000, NS32K, ROMP, Stormy16, V850, Xtensa, AVR32
 - Additional processors independently supported: D10V. LatticeMico32, MeP. Motorola 6809, MicroBlaze, MSP430, Nios II and Nios, PDP-10, TIGCC (m68k variant), Z8000, PIC24/dsPIC, NEC SX architecture

Comprehensiveness of GCC 4.3.1: Size

- Pristine compiler sources (download tarball)
 - ► Lines of C code : 2122047
 - ► Lines of MD code : 245933
 - ► Lines of total code : 2367980
 - ► Total authors (approx) : 65
 - ▶ Backend directories : 35
- Generated source for i386 (input language: c)
 - ► Total lines of code: 439703
 - ▶ Total lines of .c files code : 334855
 - ▶ Total number of .c files : 16
 - ► Total lines of h files: 104848
 - ▶ Total number of .h files : 274

Introduction: Motivation Behind this Workshop

Open Source and Free Software Development Model

 The Cathedral and the Bazaar Eric S Raymond, 1999.

Introduction: Motivation Behind this Workshop

- The Cathedral and the Bazaar Eric S Raymond, 1999.
- Cathedral: Total Centralized Control Design, implement, test, release

Open Source and Tree Software Development Woder

Introduction: Motivation Behind this Workshop

- The Cathedral and the Bazaar Eric S Raymond, 1999.
- Cathedral: Total Centralized Control Design, implement, test, release
 - ► + Facilitates a homogeneous, coherent, and well-planned design
 - ► Could be restricted by the vision of a few people

- The Cathedral and the Bazaar Eric S Raymond, 1999.
- Cathedral: Total Centralized Control Design, implement, test, release
 - ► + Facilitates a homogeneous, coherent, and well-planned design

Introduction: Motivation Behind this Workshop

- Could be restricted by the vision of a few people
- Bazaar: Total Decentralization
 Release early, release often, let users fix bugs

Open Source and Free Software Development Model

- The Cathedral and the Bazaar Eric S Raymond, 1999.
- Cathedral: Total Centralized Control Design, implement, test, release
 - ► + Facilitates a homogeneous, coherent, and well-planned design
 - ▶ Could be restricted by the vision of a few people
- Bazaar: Total Decentralization
 Release early, release often, let users fix bugs
 - ► Could lead to heterogeneity in the design
 - ► + Is not restricted by the vision of a few people

Introduction: Motivation Behind this Workshop

Release early, release often, let users fix bugs

• Brooks' law (The Mythical Man Month, 1975)

A combination of the two seems more sensible

July 09

Release early, release often, let users fix bugs

- Brooks' law (The Mythical Man Month, 1975)
 - ▶ 12 man month effort

A combination of the two seems more sensible

Introduction: Motivation Behind this Workshop

July 09

July 09

The Bazaar Approaci

Release early, release often, let users fix bugs

- Brooks' law (The Mythical Man Month, 1975)
 - 12 man month effort
 - 1 person working for 12 months OR 12 persons working for 1 month?

A combination of the two seems more sensible

The Bazaar Approach

Release early, release often, let users fix bugs

- Brooks' law (The Mythical Man Month, 1975)
 - 12 man month effort
 - I person working for 12 months OR 12 persons working for 1 month?
- Bazaar approach believes that the two somewhat equivalent in internet-based distributed development.

A combination of the two seems more sensible

The Bazaar Approach

Release early, release often, let users fix bugs

- Brooks' law (The Mythical Man Month, 1975)
 - ▶ 12 man month effort
 - 1 person working for 12 months OR 12 persons working for 1 month?
- Bazaar approach believes that the two somewhat equivalent in internet-based distributed development.
- "Given enough eyeballs, all bugs are shallow".
 Code errors, logical errors, and architectural errors.

A combination of the two seems more sensible

The Current Development Model of GCC

- GCC Steering Committee: Free Software Foundation has given charge
 - Major policy decisions
 - Handling Administrative and Political issues
- Release Managers:
 - Coordination of releases
- Maintainers:
 - Usually area/branch/module specific
 - ▶ Responsible for design and implementation

The Current Development Model of GCC

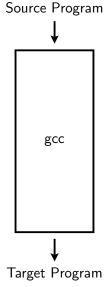
- Proposing changes
 - Extensive discussions over mailing lists
 - Submissions to gcc-patches@gcc.gnu.org
 - ► Major changes are forked off as an independent branch which is later merged with the main code
- Reviwers:
 - ► Can be general/global or area/branch/module specific
 - Can approve changes suggested by others
 - ▶ Need approval of other reviewers for their own changes
- Maintainers:
 - ► Can approve changes suggested by others
 - Do not need approval of reviewers for their own changes in their area/branch/module

Why is Understanding GCC Difficult?

Deeper reason: GCC is not a *compiler* but a *compiler generation* framework

There are two distinct gaps that need to be bridged:

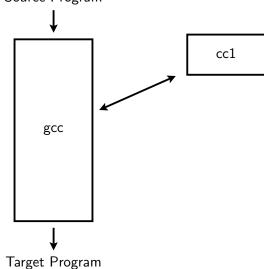
- Input-output of the generation framework: The target specification and the generated compiler
- Input-output of the generated compiler: A source program and the generated assembly program



IT Bombay

Source Program

Introduction: Motivation Behind this Workshop



July 09

cpp

20/29

Source Program

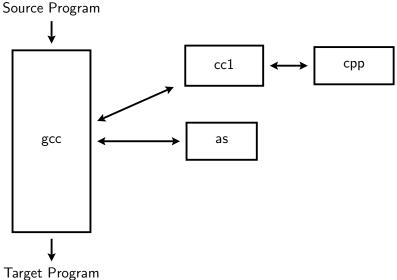
gcc

Target Program

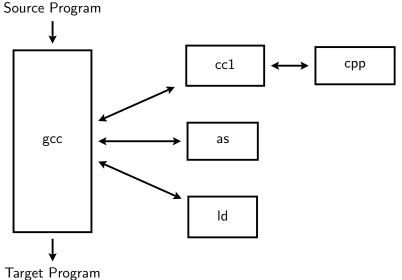
cc1

GCC Resource Center, IIT Bombay

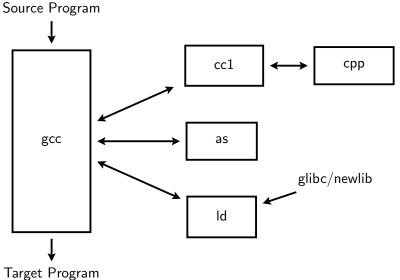
July 09



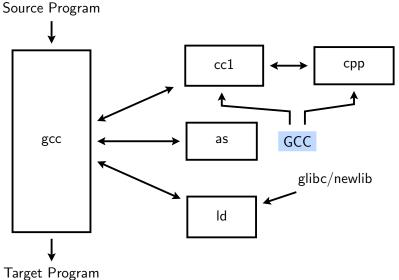
The Gnu Tool Chain



The Gnu Tool Chain



The Gnu Tool Chain



The Architecture of GCC

Compiler Generation Framework

Language Specific Code Language and Machine Independent Generic Code

Dependent Generator Code

Machine

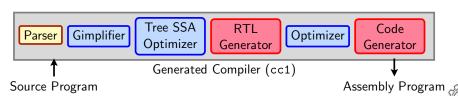
Machine Descriptions

Language and Machine Dependent Generic Code

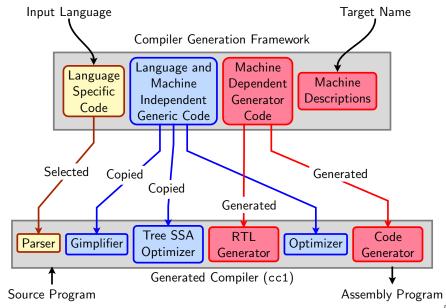
Language and Machine Dependent Generator Code

Machine Descriptions

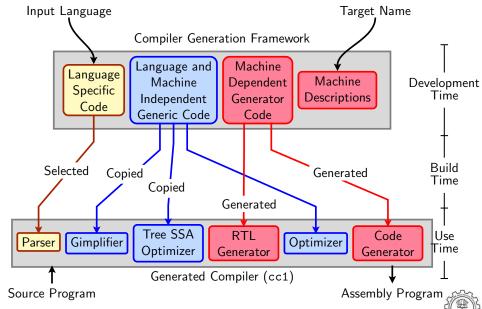
Compiler Generation Framework



The Architecture of GCC



The Architecture of GCC



Part 3

Our Philosophy and Pedogogy

Introduction: Our Philosophy and Pedogogy

July 09

22/29

gogy

Twin goals of this workshop:

- Learning how to learn GCC
- Striking a balance between theory and practice



We will

You will

July 09

Introduction: Our Philosophy and Pedogogy

Philosophy and Pedogogy

23/29

Essential Abstrations in GCC

Introduction: Our Philosophy and Pedogogy

We will

July 09

Explain configuration and building of GCC

You will

July 09

- We will
 - Explain configuration and building of GCC
 - Explain essential abstractions related to compilation
 The key intermediate representations and their manipulations

You will

- We will
 - Explain configuration and building of GCC
 - Explain essential abstractions related to compilation The key intermediate representations and their manipulations
 - Explain essential abstractions related to generation of a compiler The machine descriptions and their influence on compilation
- You will

- We will
 - Explain configuration and building of GCC
 - Explain essential abstractions related to compilation
 The key intermediate representations and their manipulations
 - Explain essential abstractions related to generation of a compiler
 The machine descriptions and their influence on compilation
- You will
 - Build and run GCC

We will

- Explain configuration and building of GCC
- Explain essential abstractions related to compilation
 The key intermediate representations and their manipulations
- ► Explain essential abstractions related to generation of a compiler The machine descriptions and their influence on compilation
- You will
 - Build and run GCC
 - Examine various IR dumps produced by GCC

We will

- Explain configuration and building of GCC
- Explain essential abstractions related to compilation
 The key intermediate representations and their manipulations
- ► Explain essential abstractions related to generation of a compiler The machine descriptions and their influence on compilation

You will

- Build and run GCC
- Examine various IR dumps produced by GCC
- ▶ Add a new machine description and systematically enhance it

- A programmer will get a better compiler
- A compiler professional will be able to deploy and enhance GCC much more easily.
- A compiler researcher will be able to use GCC for research much better.
- A compiler teacher will be able to strike a better balance between theory and practice.
- A compiler student will be exposed to issues in real compilers.

Schedule: Day 1

09:00 to 09:30	Registration
09:30 to 10:00	Introduction and opening remarks
10:00 to 11:15	Getting started with GCC: Configuration and building
11:15 to 11:30	Tea break
11:30 to 13:00	(Lab) Configuration and building
13:00 to 14:00	Lunch
14:00 to 15:00	Gray Box Probing of GCC
	An introduction to IRs
15:00 to 15:45	Introduction to Gimple IR
15:45 to 16:00	Tea break
16:00 to 17:30	(Lab) Adding a gimple pass
19:00 to 20:00	Inspecting and debugging gcc code (Optional Session)
	(use of cscope, ctags, gdb etc.)
20:00 to 21:00	Dinner

Schedule: Day 2

09:30 to 10:15	Introduction to RTL
10:15 to 11:00	An overview of retargetability and an introduction to
	machine descriptions
11:00 to 11:15	Tea break
11:15 to 13:00	(Lab) Adding an RTL pass
13:00 to 14:00	Lunch
14:00 to 15:30	Spim machine descriptions: Level 0 and 1
15:30 to 15:45	Tea break
15:45 to 17:30	(Lab) spim machine descriptions
18:30 to 20:00	(Optional session) An Overview of research projects in
	GCC Resource Center
20:00 to 21:00	Dinner

Schedule: Day 3

09:30 to 10:10	spim machine descriptions levels 2 and 3
10:15 to 10:45	
10:15 to 10:45	The retargetability model of GCC
10:45 to 11:00	Tea break
11:00 to 13:00	(Lab) spim machine descriptions
	T
13:00 to 14:00	Lunch
13:00 to 14:00 14:00 to 15:30	Lunch The Generic Data Flow Analyzer in GCC

- Lab arrangements:
 - Location: Block A, first room on the left
 - Common local account on each machine: Use the same machine for all lab assignment.
 - ▶ Login id: gccworkshop09, Password: workshop

- Lab arrangements:
 - ▶ Location: Block A, first room on the left
 - Common local account on each machine:
 Use the same machine for all lab assignment.
 - ▶ Login id: gccworkshop09, Password: workshop
- Accessing internet
 - Set proxy to proxy.cse.iitb.ac.in
 - Set no proxy for .iitb.ac.in
 - Login id: workshop, Password: workshop123

- Lab arrangements:
 - Location: Block A, first room on the left
 - Common local account on each machine: Use the same machine for all lab assignment.
 - ▶ Login id: gccworkshop09, Password: workshop
- Accessing internet
 - Set proxy to proxy.cse.iitb.ac.in
 - Set no proxy for .iitb.ac.in
 - Login id: workshop, Password: workshop123
- Tiffin/Tea at 5:30
- Dinner and breakfast arrangments
 - Breakfast available in the hostels
 - Dinner: Fri, Sat: FCKA Foyer
 - ▶ Dinner Sun, Breakfast Mon: in hostels on prior intimation

- Certificates of participation
 Can be collected from Nisha on Sunday afternoon
- Receipts of payments
 Will be sent to your address by post



July 09

- Certificates of participation
 Can be collected from Nisha on Sunday afternoon
- Receipts of payments
 Will be sent to your address by post
- Concluding session
 - Informal discussions
 - Brief (10 minute) presentations by participating organizations/individuals
 - Confirmed presentations from KPIT Cummins, and Acme Technologies
 - If you are interested, please contact me well today or latest tomorrow
 - Feedback forms