RISC 2017: Research and Innovation Symposium in Computing

Abstract Booklet

Sprint Thesis Talks

And

Poster + Demo Presentations

Department of Computer Science and Engineering

IIT Bombay

1st April, 2017

Sprint Thesis Talks

1. **Title**: Large Audience Environments: Mitigating Medium Contention in WiFi
   Collaborators: Muhammad Inamullah, Prof. Bhaskaran Raman

   Large audience environments (LAE) are special WiFi scenarios like classrooms and stadiums. LAEs are interesting in that many clients simultaneously need to download some content, while at the same time uplink traffic continues from clients to the AP. WiFi networks in general experience many performance issues due to reasons like interference from WiFi and non-WiFi sources, low signal strength, high overhead, and local contention. In addition to the general performance issues of WiFi, the LAEs' special scenario of uplink traffic and WiFi's management data contending with the more preferred download traffic poses an interesting research challenge: "How to provide a better download experience in LAEs?"

   Prior works have provided solutions like prioritizing AP traffic, and group-scheduling of downlink flows. The effect of scale on prioritized downloads, and the ways of mitigating the effect of uplink protocol data on group-scheduling need to be studied in those works respectively. More importantly, we propose to study the effect of prioritizing downlink traffic when group scheduling is employed

2. **Title**: Verification of timed critical system
   Collaborators: Sparsa Roychowdhury, Krishna S, S Akshay

   Verification of timed critical system is very important. Modelling timed systems using timed automata is a well-accepted procedure, but timed automata cannot model complex timed systems like timed systems with recursion. There are other models like timed push down systems, timed multi push down systems which can model more complex systems but showing that they have decidable emptiness problem is not as easy as timed automata and they all demand different procedures. In a recent work published in CONCUR'16, a new procedure is described which can prove decidable emptiness problem of various timed systems by capturing their behaviour as graphs and by applying the Courcelle's theorem by showing MSO definability and bounded tree-width for those class of graphs. Thus reducing the problem to the emptiness problem of tree automata which is known to be decidable. Here we extended this procedure to show decidable emptiness problem for multi push down systems with constraints like, bounded scope, bounded order and bounded phase.

3. **Title**: Venter: A complaint management system
   Collaborators: Vihari Piratla, Sunita Sarawagi

   With more than a billion population and rapidly developing economy like India, it is essential that the complaints from citizens are heard, addressed and organized effectively.

   The existing portals for complaint filing are unorganized, rely on age-old technologies like emails and token systems, require human attention, and are fragmented. Imagine several people complaining about the same issue and a human needs to manually attend them all individually. What if we can detect them all to be the same? When citizens in a locality with similar issues get together and form communities, it is possible to resolve issues in the most democratic way possible.

We wish to organize better what Indians are best at: complaints. Here at the Computer Science Department of IITB, we are developing a Complaint Management System, called Venter, leveraging various Machine Learning techniques. Venter is a community centric platform for filing issues and has the following prime objectives: (a) (far) better organization of complaints, their resolution and re-addressal (b) build communities of citizens that share a cause.

4. **Title**: Techniques for Precise and Scalable Data Flow Analysis
   Collaborators: Komal Pathade, Uday Khedker

The area of program verification concerns with assessing the correctness of the programs. Program verification needs code level information which is computed using a technique called Data Flow Analysis. Current data flow analysis techniques exhibit a tradeoff between scalability and precision i.e. techniques which can precisely gather information, do not scale to large programs due to limited computing resources while, other techniques which are scalable to large programs do not compute the information precisely. In verification of real world code, both precision and scalability are equally important. Towards this end, in this research, we propose to find techniques which will achieve a balance between scalable and precise data flow analysis.

Current scalable data flow analysis solutions called as maximal fix point solutions are imprecise because they do not distinguish the information arrived from infeasible program paths. We propose a technique to enrich existing maximal fix point solutions to filter out information arrived from infeasible program paths.

5. **Title**: Quadcopter based Unknown Planar Structure Exploration
   Collaborators: Praneeth A S, Prof. Sharat

Ever wanted to be an invisible, silent wasp and see 'what's out there' but bemoaned your lack of wings? Ever wanted to work for the country's intelligence and explore hidden caves in mountainous regions far from your cozy, warm office? More mundane, did you want to see what you have stored in your home attic many eons ago when you are trying to find your long missing suitcase? Never fear, I am there for you.

In this work, we show how a quadcopter can be used to explore unknown structures. For the purpose of simplification, we assume an isolated region that contains a structure made up of multiple planar surfaces (such as a temple or a stupa). Our algorithm dictates commands to a camera empowered drone that autonomously navigates around the structure, and capture pictures of those portions of the planes that a user proclaims interest. To recall, these surfaces happen to be arranged such that all surfaces are not visible from any single viewpoint --- so your smartphone won't cut it.

You can take these pictures and do what you want -- we suggest a panoramic mosaic.
And oh, our drone is assumed to be wallet-friendly, programmable, and *not* have a GPS (so you can use it indoors).

6. **Title**: Multilevel Differentiated Hypervisor Caching for Derivative Clouds
   Collaborators: Prashanth, Purushottam Kulkarni

Derivative clouds, light weight application containers provisioned in virtual machines, are increasingly becoming popular cost-effective alternatives for IaaS and PaaS services. However, in such nested virtualization setups used in the context of derivative clouds, hypervisors and application containers are not designed to integrate in a seamless manner which causes inefficiencies in satisfying cloud provider SLAs. Hypervisor caching is a resource management technique used by cloud providers to enhance storage efficiency which in turn boosts application performance. However, hypervisor caching provides no support for multi-level differentiated provisioning, and therefore is not directly adaptable in derivative clouds. We propose DoubleDecker, a hypervisor caching framework which provides two cache management interfaces, memory and SSD for differentiated cache partitioning and management in nested virtualization setups. By extending the application container level relative weights to the hypervisor cache, DoubleDecker can improve the application performance by up to 11 times. As part of our current work, we are working on providing a hybrid mode for cache partitioning where caches can be configured at both levels and providers could provide SLAs that would map to appropriate cache partitions.

7. **Title**: Question Generation Using Knowledge base and Text
   Collaborators: Vishwajeet kumar, Ganesh Ramakrishnan

We present a system of question generation by projecting knowledge base (e.g. wikidata) onto Text. The main goal is to engage learners and test factual knowledge of a learner. Our approach projects the knowledge base onto the corpus to annotate entities and relationship present in the text to extract existing facts. Some examples of factual questions are "**When did Alexander invade India?**", "**Who invented small pox vaccine?**". It uses a encoder to encode lexical tokens and facts into a n-dimensional continuous vector representation. The encoded fact and context is decoded using LSTM-RNN model to generate question. Various other aspects of the system are a) generating questions based on hardness criteria b) generate a set of distractions as multiple choices. We also plan to provide an interface to a teacher to judge the usefulness of a question

8. **Title**: MPTCP on smartphones
   Collaborators: Avinash Kumar Chaurasia, Bhaskar Raman

TCP is one of the most prominent transport layer protocol used in today's Internet. Together with UDP it makes the Internet what it is today. In early days of Internet, most systems had single network interface. However, now many devices such as multihomed servers, laptops etc that connect to internet have multiple interfaces. In recent years with the advent of smartphones more and more devices are coming on internet with multiple interfaces i.e. WiFi and 3G/4G. Architects of TCP never thought of using multiple interfaces while they designed the TCP. User's Internet experience did improve with the help of multiple interfaces as when network over one interface is down, it can use another. But, there is still a room to use these multiple interfaces efficiently. Multipath TCP (MPTCP) intends to use these multiple interfaces simultaneously and efficiently. None of the current smartphone carry MPTCP, this makes it tougher to understand the behavior of protocol and it's usefulness with respect to smartphone user. However, lots of studies has been done for static clients (smartphones, desktop, etc.), most of the bad experiences related to internet comes when user is actually mobile. My current work is on understanding the nature of this protocol while the person is mobile. To see whether MPTCP can really benefits the user experience on smartphones.

9. **Title**: Multi-resolution bank convolutional neural network for learning the keypoint descriptors
Collaborators: Rahul Mitra, Sharat Chandran

In this project, we use multi-resolution bank convolutional neural network for learning the keypoint descriptors. We also propose the Photosynth (PS) dataset consisting of substantially more number of scenes, images, and positive and negative correspondences compared to the Multi-View Stereo (MVS) dataset. The PS dataset also has better coverage of the overall viewpoint, scale, and lighting challenges in comparison to the MVS dataset. We evaluate our approach on two standard datasets, viz. Oxford Affine Covariant Regions Dataset (ACRD) to quantify the image descriptor performance. We report, improved performance of ConvNets trained on MVS dataset. We also reconstruct 3D point clouds from images provided in the Strecha dataset by finding point-correspondences using our descriptor and SIFT. We observed that our reconstructions had more points and less moise compared to SIFT.

10. **Title**: Distributed Safety Games: The Two Process Case
Collaborators: Nehul Jain, Bharat Adsul

Safety games played on transition systems have been well studied and serve as a basis for the problem of automatically synthesizing programs realizing the given specification. We propose a new model of distributed safety games where a team of distributed agents/processes compete against the environment to achieve the given safety objective. The environment schedules and chooses 'local' actions as in a sequential game. However, the relevant processes react based only on their 'casual past'. On a synchronization, the participating processes are allowed to exchange   all   their   past   information.

The central algorithmic problem is to decide if there exists a distributed winning strategy in the given distributed safety game. We believe this to be decidable. We carry out an in-depth analysis of these games in the two process case. Our results include
a)   an   exponential-time   algorithm   for   solving   these   games,
b) a hardness result showing that solving these games is NP-complete and
c) a novel investigation of the issues related to the 'determinacy' of these games.

11. **Title**: Venter: A Complaint management system
Collaborators: Vihari Piratla, Sunita Sarawagi

With more than a billion population and rapidly developing economy like India, it is essential that the complaints from citizens are heard, addressed and organized effectively. The existing portals for complaint filing are unorganized, rely on age-old technologies like emails and token systems, require human attention, and are fragmented. Imagine several people complaining about the same issue and a human needs to manually attend them all individually. What if we can detect them all to be the same? When citizens in a locality with similar issues get together and form communities, it is possible to resolve issues in the most democratic way possible.

We wish to organize better what Indians are best at: complaints. Here at the Computer Science Department of IITB, we are developing a Complaint Management System, called Venter, leveraging various Machine Learning techniques. Venter is a community centric platform for filing issues and has the following prime objectives: (a) (far) better organization of complaints, their resolution and re-addressal (b) build communities of citizens that share a cause.

12. **Title**: Analysis of reconstruction errors for compressed sensing under Poisson noise
Collaborators: Deepak Garg, Ajit Rajwade

The analysis of reconstruction errors for compressed sensing under Poisson noise is challenging due to the signal dependent nature of the noise, and also because the Poisson negative log-likelihood is not a metric. In this paper, we present error bounds for reconstruction of signals which are sparse or compressible under any given orthonormal basis, given compressed measurements corrupted by Poisson noise and acquired in a realistic physical system. The concerned optimization problem is framed based on the well-known Variance Stabilization Transforms which transform the noise to (approximately) Gaussian with a fixed variance. This problem also turns out to be convex. We demonstrate promising numerical results on signals with different sparsity, intensity levels and given different numbers of compressed measurements.

13. **Title**: Investigations into subword units for statistical machine translation between related languages
Collaborators: Anoop Kunchukuttan, Pushpak Bhattacharya

Related languages are those that exhibit lexical and structural similarities on account of sharing a common ancestry or being in contact for a long period of time. Machine Translation between related languages is a major requirement since there is substantial government, commercial and cultural communication among people speaking related languages. However, most of these languages have few parallel corpora resources, an important requirement for building good quality statistical machine translation (SMT) systems.

A key property of related languages is lexical similarity, which means the languages share many words with the similar form (spelling/pronunciation) and meaning. These words could be cognates, lateral borrowings or loan words from other languages. Modelling lexical similarity among related languages is the key to building good-quality SMT systems with limited parallel corpora. We propose the use of two subword units of translation for modelling lexical similarity: (i) orthographic syllables motivated from the design of Indic scripts and (ii) byte pair encoded units inspired from compression theory. We show that the proposed significantly outperforms other units of representation (word, morpheme and character), over multiple language pairs, spanning different language families, with varying degrees of lexical similarity and is robust to domain changes too.

14. **Title**: Language-based Security and Jif programming experiences
Collaborators: Sandip Ghosal, G. Sivakumar, R.K Shyamasundar

Information flow control among mutual distrusted authorities inside a program provide fine-grained application level security and thus reduce the burden of securing a whole system. Information flow specifications are provided by the users through different annotations in a high-level programming language. The compiler of that particular language certify the application if it is information flow-safe. An application is called secure in terms of information flow if under any circumstances there exist no information flow between two distrusted agents. The Jif package [1] provides such a compiler which checks the information flow and announce a program secure if it conforms to information flow rules, those specified using the Jif annotations in a Java program. Since last one decade Jif language has been used extensively in the field of information flow control in language-based security. In this report we have presented an assessment of Jif language and discussed various shortcomings with examples and

experiment results. Advantage of dynamic labeling, possible vulnerability in non-robust and robust declassification, information flow through covert channels and experiment with subject labels are the few aspects of information flow control that we will discuss with respect of Jif language.

15. **Title**: Challenges in Software-defined networking (SDN)
    Collaborators: Rinku Shah, Mythili Vutukuru, Purushottam Kulkarni

    Software-defined networking, aka, SDN, along with benefits, has multiple challenges like Performance & Scalability, Security, Reliability, and Fault-tolerance. We are addressing the Performance & Scalability challenge. We define three bottleneck reasons in a SDN-based application design- Controller, Data-plane switches, and Switch-Controller communication. Existing literature broadly classifies the scalability designs as- Centralized & Multi-threaded, and Distributed(Horizontal & Vertical) Controller designs.

    We consider the Vertically-distributed controller design, where the application-functions are distributed among the central-controller and the local-controller, based on the network-visibility requirement, and locality. We define different vertical-distribution designs, and provide empirical comparison, using the use-case of LTE- EPC. The task of partitioning the tasks for local-controller, and central-controller is currently done by the application programmer. We propose automation of application partitioning, placement, and configuration, using the hints given by the programmer. After looking at the variety of use-cases, if we find that the input traffic-mix dictates the choice of design, we would provide the ability to our framework, to reconfigure the controller architecture, as per the current traffic-mix, to satisfy the application's performance & scale requirements. One of our designs(Design-1) shows an improvement of 21.8% for Service-Request, & 13.5% improvement for Attach-Detach procedures. The second design(Design-2) shows an improvement of 24% for Service-Request. The third design(Design 4) shows an improvement of 17% for ServiceRequest, and degrades the Attach-Detach procedure by 9%

16. **Title**: Optimization of Multi Village Piped Water Schemes
    Collaborators: Nikhil Hooda, Om P. Damani

    The classic problem of the capital cost optimization of branched piped networks consists of choosing pipe diameters for each link in the network from a discrete set of commercially available pipe diameters. Each link in the network can consist of multiple pipe segments of differing diameters. Water networks also consist of intermediate tanks that act as buffers between incoming flow from the primary source and the outgoing flow to the demand nodes. The network from the primary source to the tanks is called primary network, and the network from the tanks to the demand nodes is called secondary network. During the design stage, the primary and secondary networks are typically optimized separately with the tanks acting as demand nodes for the primary network. The choice of tank locations, their elevations, and the set of demand nodes to be served by different tanks, is manually made in an adhoc fashion before any optimization is done. It is desirable therefore to include this tank configuration choice in the pipe diameter optimization process itself. We have constructed an Integer Linear Program (ILP) model that optimally solves the problem of tank configuration and pipe diameter selection. This model has further been extended to include other network components like pumps and valves. The model has been tested on several real life networks from Thane district in Maharashtra, India. The model has been

implemented as a freely available web application called JalTantra. It is being used by several Rural Water Supply Department engineers in Maharashtra, India.

17. **Title**: Identifying a "good arm" in a Large Multi-Armed Bandit
    Collaborators: Arghya Roy Chaudhuri, Prof. Shivaram Kalyanakrishnan

Sequential Decision Making (SDM) under uncertainty is a p opular area in the field of Reinforcement Le arning (RL). Multi-Armed Bandit(MAB) is a branch of SDM in which the ob jective is to output a strategy that maximizes the gain for a given set of slot machines. Considerable amount of research has b een done on settings like Regret Minimization, PAC etc. The applicability of MAB ranges from Online advertisement, Drug tes ting, Automated game playing to scenarios where exploration and exploitation trade-off plays the main role in optimizing the objective. Our current research interest deals with large-scale MAB problems. In case of real life problems, quite often the decision maker is happy with quickly identifying a "good enough" decision. Our approach fulfills this exact requirement for fast and good decision making. We have proposed a PAC formulation for identifying an arm in an n-armed bandit whose mean is within a fixed tolerance of the m-th highest mean. The key implication of our proposed approach is the ability to derive upper bounds on the sample complexity that depends on n/m in place of n. Consequently, even when the number of arms is infinite, we only need a finite number of samples to identify an arm that compares favorably with a fixed reward quantile. Hence our approach is attractive to the above applications, wherein the number of arms (ad-parameters, molecular configurations, playing strategies etc.) may run into a few thousand or more.

18. **Title**: Dynamic Server Consolidation for cyclic workload
    Collaborators: Anshu Yadav, Varsha Apte

The virtualization technology allows applications to run inside a Virtual Machine (VM) and then hosting multiple VMs on a single Physical Machine (PM) in an isolated manner. The load on these applications, and hence, the resource requirement of the VMs varies with time. Therefore, a one-time static VM-PM mapping may lead to under or over utilization of physical resources. The Dynamic Server Consolidation Problem (DSCP) refers to finding a time varying (dynamic) mapping of VMs on PMs as per the change in resource requirement of the VMs. The dynamic mapping of VMs on PMs requires live VM migration, which is a costly operation, as it negatively affects the performance of the applications running inside the migrating as well as the co-located VMs. Hence the goal of DSCP is to minimize the number of active PMs as well as the number of migrations. Lots of work has been done in this area, and a number of heuristic based solutions have been proposed for DSCP, but most of them are reactive, in the sense that the VM-PM mapping decisions are taken online based on the current mapping and the current resource requirement of the VMs.

However, it has been found that the demand pattern of most of the web-based applications follow periodic trend, which implies that the resource requirement of the VMs, for an entire cycle can be predicted beforehand, based on the previous resource requirements. Under this assumption, one should be able to exploit the information of the future resource requirement of the VMs to come up with a more stable VM-PM mapping as compared to the reactive solutions. As per our knowledge, this assumption of the knowledge of the future workload has not been much explored to design a solution for dynamic server consolidation. However, such an assumption of cyclic resource demand pattern and hence, the knowledge of future load has been

made for other problem definitions, like static VM-PM mapping, resource allocation problem for VMs, etc.

Our research problem focuses on proposing a stability based solution to find dynamic VM-PM mapping under the assumption of cyclic workload, which obviously implies the knowledge of future resource requirement of the VMs. Thus our goal is to optimize the long term profit/cost of running a data center with multiple VMs and PMs by avoiding unnecessary VM migrations. The cost being defined as some function of the cost of PMs, cost of migrations, etc.

19. **Title**: Generalized Points-to Graphs: A New Abstraction of Memory in Presence of Pointers
Collaborators: Pritam Gharat, Uday Khedker

Computing precise (fully flow-sensitive and context-sensitive) and exhaustive (as against demand-driven) points-to information is known to be computationally expensive. Therefore many practical tools approximate the points-to information trading precision for efficiency. This often has adverse impact on computationally intensive analyses such as model checking. Prior approaches to flow- and context-sensitive points-to analysis (FCPA) have not scaled; for top-down approaches the problem centres on repeated analysis of the same procedure, while for bottom-up approaches the abstractions used to represent procedure summaries have proved not to scale while preserving precision. A bottom-up approach for points-to analysis requires modelling unknown pointees accessed indirectly through pointers. Such accesses are commonly handled by using placeholders to explicate unknown pointees or by using multiple call-specific summary flow functions. We propose a novel abstraction called the Generalised Points-to Graph (GPG) which generalizes the concept of points-to relations using the counts of indirection levels leaving the unknown pointees implicit. This allows us to construct GPGs without the need of placeholders. The central point is that the GPG not only compactly represents procedure summaries (in terms of a compact representation for memory, memory updates and the control flow between these updates), but also can be efficiently and compactly inlined when one procedure calls another. GPGs allow performing FCPA by progressively reducing the counts of indirection levels of points-to relations via data-flow analysis. GPGs distinguish between *may* and *must* pointer updates thereby facilitating strong updates within calling contexts. At a practical level, GPGs hold a promise of efficiency and scalability for FCPA without compromising precision. At a more general level, GPGs provide a convenient abstraction of memory in presence of pointers. Static analyses that are influenced by pointers may be able to use GPGs by combining them with their original abstractions.

20. **Title**: Verifying Array Manipulating Programs by Tiling
Collaborators: Divyesh Unadkat, Supratik Chakraborty and Ashutosh Gupta

Formally verifying properties of programs manipulating arrays is computationally challenging. State-of-the-art tools often resort to reasoning about pre-defined classes of array access patterns in loops to prove properties of such programs. However, several common array access patterns do not belong to these pre-defined classes. Moreover, programs with new array access patterns continue to be written. In this paper, we present a novel property-
driven verification method that first infers array access patterns in loops using a simple heuristic, and then uses this information to compositionally prove universally quantified assertions about arrays. Specifically, we partition array accesses in a loop into tiles, such that the following hold: (i) eachiteration of the loop modifies array elements only within a designated tile,(ii) the collection of tiles covers the range of all array elements relevant to the property of interest, and (iii) a loop iteration updates its associated tile in a manner such that

the desired property holds on all array elements in the tile. These conditions can be easily checked using SMT solvers, and their conjunction implies that the property of interest holds on all array elements of interest after the loop terminates. We have implemented our method in a tool called Tiler. Initial experiments show that Tiler outperforms Vaphor, Booster, BMC, and SMACK+Corral on several interesting benchmarks.

21. **Title**: Transductions and its applications
    Collaborators: Vrunda Dave, Krishna S

    Our work is in the area of transductions and its applications in the area of formal verification. A transduction can b e thought of as a program which computes some function/relation based on a set of inputs. It can therefore be used as a mo del for a program. Being able to answer questions related to transducers can help us identify properties of programs. For example, program equivalence reduces to checking equality of transducers. Transducers are instrumental in infinite state model checking, and automatic speech recognition to name a few applications. There are several formalisms for defining transductions, e.g., Logic, two way transducers (2WST), streaming string transducers (SST) etc. We restrict ourselves to transductions over finite and infinite strings. Like in language case, there are known equivalences between these formalisms. These results, along with connection to algebraic theory, has created a robust theory of transducers. However, several areas of this topic still remain unexplored. Our recent research focused on First Order Logic(FO) definable transformations of infinite strings. In particular, we showed the equivalence of aperiodic 2WST with star-free lookaround, aperiodic SST and FO transducer for infinite strings(FSTTCS'16). We are currently working to get the algebraic characterization of transducers.

22. **Title**: Post-processing OCR output.
    Collaborators: Rohit Saluja, Prof. Ganesh and Prof. Parag

    Did you knew that any off the shelf Vocabulary is always incomplete for curating the text in Indian languages? The OCR accuracy for documents in practice is much below English. In such a scenario detecting the erroneous words and correcting them is challenging for machines as well as humans. Deep neural network based memory models are very helpful in this scenario as they can learn OCR specific error patterns as well as language. We will present the basic models that helped in our experiments.

23. **Title**: Cache-Based Side Channel Attack on AES
    Collaborators: Ashokkumar C, Bernard Menezes , G. Sivakumar

    The software implementation of AES is an especially attractive target for cache-based side channel attacks on AES since it makes extensive use of cache-resident table look-ups. Modern processors employ hardware prefetching to reduce memory latency (cache lines are fetched in anticipation of their future use). This greatly complicates access-driven attacks since they are unable to distinguish between a line fetched on demand versus one prefetched and not subsequently used during a run of a victim running AES. Unlike previous work which does not explicitly handle hardware prefetching, our multi-threaded spy code and key retrieval algorithms are designed to succeed even in the presence of prefetching albeit at the cost of requiring more blocks of ciphertext. We demonstrate through implementations on real machines corroborated by analytical models that, with probability 95%, we are able to recover

the AES key using 25 blocks of ciphertext in the presence of prefetching and, stunningly, a mere 3-5 blocks with prefetching disabled. Moreover, our implementation is error-tolerant and also succeeds on the i3/i5/i7 processors which are equipped with highly aggressive prefetchers.

24. **Title**: Mitigating nesting-agnostic hypervisor policies in derivative clouds.
Collaborators: **Chandra Prakash**, Prashanth, Purushottam Kulkarni, Umesh Bellur

The fixed granularity of virtual machines offered by IaaS providers has prompted the evolution of derivative clouds. With a derivative setup where containers are provisioned within virtual machines, the guest OS manages virtual resources inside a VM whereas the hypervisor manages the physical resources distributed among VMs. This results in two control centers over the set of resources used by containers. The hypervisor takes control actions such as memory ballooning or the withdrawal of a virtual CPU to manage over-provisioning without being aware of the effect these actions will have on individual containers inside the VM. The derivative cloud provider executing containers in the VM needs a mechanism to react to such changes — based on resource management policies setup for this purpose.

25. **Title:** GPU-assisted rapid memory deduplication in virtualization environments
Collaborators: **Anshuj Garg**, Debadatta Mishra, Purushottam Kulkarni

Content based page sharing techniques improve memory efficiency in virtualized systems by identifying and merging identical pages. Kernel Same-page Merging (KSM), a Linux kernel utility for page sharing, sequentially scans memory pages of virtual machines to deduplicate pages. Sequential scanning of pages has several undesirable side effects— wasted CPU cycles when no sharing opportunities exist, and rate of discovery of sharing being dependent on the scanning rate and corresponding CPU availability. In this work, we exploit presence of GPUs on modern systems to enable rapid memory sharing through targeted scanning of pages.

Our solution works in two phases, the first where pages of virtual machines are processed by the GPU to identify likely pages for sharing and a second phase that performs page-level similarity checks on a targeted set of sharable pages. Opportunistic usage of the GPU to produce sharing hints enables rapid and low-overhead duplicate detection, and sharing of memory pages in virtualization environments. Our Solution can achieve higher memory sharing in lesser time compared to different scan rate configurations of KSM, at lower or comparable compute costs.

Poster + Demo Presentations

1. **Title**: Hierarchical Generative Modeling and Monte-Carlo EM in Riemannian Shape Space for Hypothesis Testing
   Collaborators: Saurabh Shigwan, Suyash Awate

   Statistical shape analysis has relied on various models, each with its strengths and limitations. For multigroup analyses, while typical methods pool data to fit a single statistical model, partial pooling through hierarchical modelling can be superior. For *pointset shape* representations, we propose a novel *hierarchical model* in *Riemannian shape space*. The inference treats individual shapes and group-mean shapes as latent variables, and uses expectation maximization that relies on sampling shapes. Our generative model, including shape-smoothness priors, can be robust to segmentation errors, producing more compact per-group models and realistic shape samples. We propose a method for efficient sampling in Riemannian shape space. The results show the benefits of our hierarchical Riemannian generative model for hypothesis testing, over the state of the art.

2. **Title**: Towards Lower Bounds on Number of Dimensions for Word Embeddings
   Collaborators: Kevin Patel, Pushpak Bhattacharya

   Word embeddings are a relatively new addition to the modern NLP researcher's toolkit. However, unlike other tools, word embeddings are used in a black box manner. There are very few studies regarding various hyperparameters. One such hyperparameter is the dimension of word embeddings. They are rather decided based on a rule of thumb: in the range 50 to 300. In this paper, we show that the dimension should instead be chosen based on corpus statistics. More specifically, we show that the number of pairwise equidistant words of the corpus vocabulary gives a lower bound on the number of dimensions, and going below this bound results in degradation of learned word embeddings. Through our evaluations on standard word embedding evaluation tasks, we show that for dimensions higher than or equal to the bound, we get better results as compared to the ones below it.

3. **Title**: Noun Compound Interpretation
   Collaborators: Girishkumar Ponkiya, Pushpak Bhattacharya

   Noun compounds (or noun sequences) are a productive, continuous sequence of more than one nouns. Most of the noun compounds appear only once in a large corpus. These characteristics of noun compounds make them a special case, and demand special treatment. Here, the problem is to find noun compounds from text, parse them if required, and extract semantic relation between components of the noun compound. A task of extracting an abstract relation between components of the noun compound (e.g., apple pie: Made Of), or paraphrasing noun compound using verb and prepositions (apple pie : "a pie made of apple" or "a pie with apple flavor"), is known as interpretation of noun compound (or noun compound inter-pretation). For our work, we use a set of predefined abstract labels as semantic relations.

   Following are major bottlenecks in current system, and our approaches to solve the same:

1. There is no acceptable inventory of semantic relations. We have analyzed a inventory of semantic relations [1], and we trying to refine the inventory. We are also planning to use a data driven approach which will help in refining the current inventories.

2. Inspite of millions of noun compounds in large corpora, there is no sufficiently large annotated dataset for supervised training. We are planning to use semi-supervised approach to tackle this.

3. Context influences semantic relations. But, the present datasets have annotation for each noun compound without context. We are planning to study how context can be used for the task. We are planning to use web-extracted information to bring the context in play.

4. **Title:** Multilingual Neural Transliteration
   Collaborators: Anoop Kunchukuttan, Pushpak Bhattacharya

We present initial results from our work on multilingual neural machine transliteration. We use an encoder-decoder architecture inspired from similar multilingual architectures proposed recently for neural machine translation, but adapt it for the transliteration task. Specifically, our adaptations are as follows: (i) use of a convolutional neural network based encoder, (ii) use of shared embeddings between languages with compatible scripts, (iii) use of richer phonetic feature based input.

We show that multilingual transliteration exhibits significant improvement in transliteration accuracy over bilingual transliteration and our proposed adaptations yield further improvements in transliteration accuracy. We also show that the multilingual transliteration models can generalize well to language pairs not encountered during training and hence perform well on the zero shot transliteration task.

5. **Title**: Partial Marking for Automated Grading of SQL Queries
   Collaborators: Bikash Chandra, S. Sudarshan

The XData system, currently being developed at IIT Bombay, provides an automated and interactive platform for grading student SQL queries, as well as for learning SQL. Prior work on the XData system focused on generating query specific test cases to catch common errors in queries. These test cases are used to check whether the student queries are correct or not. For grading student assignments, it is usually not sufficient to just check if a query is correct: if the query is incorrect, partial marks may need to be given, depending on how close the query is to being correct. We have extended the XData system by adding features that enable awarding of partial marks to incorrect student queries. Our system is able to go beyond numerous syntactic features when comparing a student query with a correct query. These features of our grading system allow the grading of SQL queries to be fully automated, and scalable to even large class sizes such as those of MOOCs.

6. **Title**: Synergistic Program Analyses (SPA)
   Collaborators: Anshuman Dhuliya, Uday Khedker

Program analysis is a central component of almost every compiler related tool. However, even after decades of work and research in program analysis, there is still an urgent need for static techniques which provide more precise information to discover program invariants. The solution may lie in combining isolated analyses together. Combined analyses can have mutually

beneficial interactions that improve precision in many cases. The traditional ways of combining analyses have focused either on manually combining few analyses together with tailored interleaved interactions, or to cascade analyses with transformations in phases. While the phases introduce the undecidable phase ordering problem, the manual method of combining analyses into a monolithic analysis provides a tailored solution to problems arising in specialized cases. Given the vast number of analyses in use today and the continuous addition of new ones, both techniques fail to provide the desired solution to improve precision in program analysis.

We propose here a semi-automated system that builds synergy between dynamically chosen analyses to give more precise results. The synergy is achieved by enabling mutually beneficial fine grained interactions between the chosen analyses. The fine grained interaction becomes possible when each analysis exposes the kind of statements it cannot resolve on its own, and the kind of data-flow information it generates. When this information is available to the system, that we call, the Synergistic Program Analyzer (SPA), it becomes possible to dynamically select the analyses required for the task in hand and provide the required synergy. The system though only intuitively defined, has shown better precision over the other known techniques.

7. **Title**: Dictionary Learning With Robust Sparsity Fitting For MRI Reconstruction
   Collaborators: Kratika Gupta, Suyash Awate

Dictionary learning is a paradigm of data adaptive representation that involves learning a set of atoms from the data such that any signal can be expressed as a linear combination of these dictionary atoms. Non negative sparse coding (NNSC) is a technique of dictionary learning that represents the input data as a non-negative sparse combination of dictionary atoms and also learns the dictionary adaptively from the data. We propose a dictionary learning framework that uses regularizing prior on the atoms and adapts to the noise model. We propose to add a smoothness prior to NNSC which helps us learn dictionary atoms that are smooth. We expect to get better fits to data by using the dictionary atoms learnt by enforcing regularity.

We use this regularized dictionary learning technique to learn dictionary atoms from medical data acquired using magnetic resonance imaging (MRI). The data acquisition method, called multishell high resolution diffusion imaging (multishell HARDI), is a time taking process as it requires multiple acquisitions. We also propose a compressed sensing framework for multishell HARDI that enables efficient reconstruction from undersampled data and consequently reduces MRI scan time.

8. **Title**: Towards Harnessing Memory Networks for Coreference Resolution
   Collaborators: Joe Cheri Ross, Pushpak Bhattacharya

Coreference resolution task demands comprehending a discourse, especially for anaphoric mentions which require semantic information for resolving antecedents. We investigate into how memory networks can be helpful for coreference resolution when posed as question answering problem. The comprehension capability of memory networks assists coreference resolution, particularly for the mentions that require semantic and context information. We experiment memory networks for coreference resolution, with 4 synthetic datasets generated for coreference resolution with varying difficulty levels. Our system's performance is compared with a traditional coreference resolution system to show why memory network can be promising for coreference resolution.

9. **Title**: Thank Goodness! A Measure for Sentence Fluency
   Collaborators: Sandeep Mathias, Pushpak Bhattacharya

   We propose a method to calculate the sentence fluency of student essays. We do this by looking at other essays in the data set. We first score individual words and phrases of the essay a property called goodness - and use those scores to predict the overall sentence fluency score of the essay. We compare our approach to a language modeling baseline, as well as a State-of-the-Art system used in scoring essays.

10. **Title**: Framework to correct mistakes in Sanskrit OCR
    Collaborators: Rohit Saluja, Prof. Ganesh and Prof. Parag

    A practical demo for Panini grammar rule based methods for correcting the OCR words in Sanskrit: the Indian language with highest number of inflections. The attendees will get on hand experience of the pain that one goes on while correcting the OCR text in Sanskrit which majorly consists of words with much larger lengths as compared to English. Such experience, which is not possible to give via a presentation, would be able to appreciate the system we have developed. We will present the graphs to publicise our framework.

11. **Title**: Verifying Array Manipulating Programs by Tiling
    Collaborators: Divyesh Unadkat, Supratik Chakraborty and Ashutosh Gupta

    Formally verifying properties of programs manipulating arrays is computationally challenging. State-of-the-art tools often resort to reasoning about pre-defined classes of array access patterns in loops to prove properties of such programs. However, several common array access patterns do not belong to these pre-defined classes. Moreover, programs with new array access patterns continue to be written. In this paper, we present a novel property-
    driven verification method that first infers array access patterns in loops using a simple heuristic, and then uses this information to compositionally prove universally quantified assertions about arrays. Specifically, we partition array accesses in a loop into tiles, such that the following hold: (i) eachiteration of the loop modifies array elements only within a designated tile,(ii) the collection of tiles covers the range of all array elements relevant to the property of interest, and (iii) a loop iteration updates its associated tile in a manner such that the desired property holds on all array elements in the tile. These conditions can be easily checked using SMT solvers, and their conjunction implies that the property of interest holds on all array elements of interest after the loop terminates. We have implemented our method in a tool called Tiler. Initial experiments show that Tiler outperforms Vaphor, Booster, BMC, and SMACK+Corral on several interesting benchmarks.

12. **Title**: Transductions and its applications
    Collaborators: Vrunda Dave, Krishna S

    Our work is in the area of transductions and its applications in the area of formal verification. A transduction can b e thought of as a program which computes some function/relation based on a set of inputs. It can therefore be used as a mo del for a program. Being able to answer questions related to transducers can help us identify properties of programs. For example, program equivalence reduces to checking equality of transducers. Transducers are instrumental in infinite state model checking, and automatic speech recognition to name a few applications. There are several formalisms for defining transductions, e.g., Logic, two way transducers (2WST), streaming string transducers (SST) etc. We restrict ourselves to transductions over

finite and infinite strings. Like in language case, there are known equivalences between these formalisms. These results, along with connection to algebraic theory, has created a robust theory of transducers. However, several areas of this topic still remain unexplored. Our recent research focused on First Order Logic(FO) definable transformations of infinite strings. In particular, we showed the equivalence of aperiodic 2WST with star-free lookaround, aperiodic SST and FO transducer for infinite strings(FSTTCS'16). We are currently working to get the algebraic characterization of transducers.

13. **Title**: Analysis of reconstruction errors for compressed sensing under Poisson noise
    Collaborators: Deepak Garg, Ajit Rajwade

    The analysis of reconstruction errors for compressed sensing under Poisson noise is challenging due to the signal dependent nature of the noise, and also because the Poisson negative log-likelihood is not a metric. In this paper, we present error bounds for reconstruction of signals which are sparse or compressible under any given orthonormal basis, given compressed measurements corrupted by Poisson noise and acquired in a realistic physical system. The concerned optimization problem is framed based on the well-known Variance Stabilization Transforms which transform the noise to (approximately) Gaussian with a fixed variance. This problem also turns out to be convex. We demonstrate promising numerical results on signals with different sparsity, intensity levels and given different numbers of compressed measurements.

14. **Title**: Venter: A complaint management system
    Collaborators: Vihari Piratla, Sunita Sarawagi

    With more than a billion population and rapidly developing economy like India, it is essential that the complaints from citizens are heard, addressed and organized effectively. The existing portals for complaint filing are unorganized, rely on age-old technologies like emails and token systems, require human attention, and are fragmented. Imagine several people complaining about the same issue and a human needs to manually attend them all individually. What if we can detect them all to be the same? When citizens in a locality with similar issues get together and form communities, it is possible to resolve issues in the most democratic way possible. We wish to organize better what Indians are best at: complaints. Here at the Computer Science Department of IITB, we are developing a Complaint Management System, called Venter, leveraging various Machine Learning techniques. Venter is a community centric platform for filing issues and has the following prime objectives: (a) (far) better organization of complaints, their resolution and re-addressal (b) build communities of citizens that share a cause.

15. **Title**: Improving IO latency of Containers by Adaptive Provisioning of Non volatile memory
    Collaborators: Shyamli Rao, Purushottam Kulkarni

    Container virtualization is an emerging technology in the area of cloud computation. It allows multiple isolated instances to share system resources such as CPU, memory, disk IO, network etc. A container runs a set of processes in an isolated environment which is allocated a bunch of resources by host operating system. The host OS should be aware of various techniques for allocating and managing the resources such as CPU, memory, storage, network etc. across all containers. Host OS has to multiplex these resources amongst the containers in an efficient way. Here we are mainly concerned with storage management. Modern data centers consolidate more workloads to reduce hardware and maintenance costs in virtualized environment. Storage performance of workloads under contention is poor due to high disk latency. Sometimes data

is stored in network attached storage (NAS), therefore every disk IO request has to go through network which increases IO latency. One of the ways to deal with this problem is provisioning a non-volatile memory as a second level cache between page cache and disk. For years, the increasing popularity of flash memory has been changing storage systems. Flash-based solid-state drives are widely used as a new cache tier on top of hard disk drives (HDDs) to speed up data-intensive applications. Caching the blocks at host side in data centers will incur better IO performance than non caching systems.