INFOLINE

VOLUME IX ISSUE II OCTOBER 2018

DEPARTMENT OF COMPUTER TECHNOLOGY AND INFORMATION TECHNOLOGY

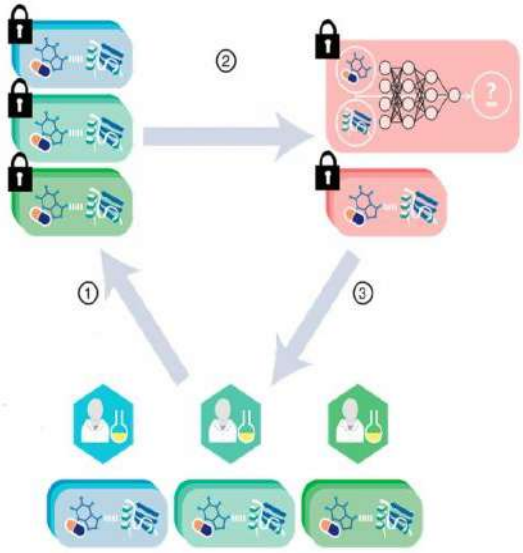
KONGU ARTS AND SCIENCE COLLEGE
(Autonomous)
Affiliated to Bharathiar University, Coimbatore.
Approved by UGC and AICTE and Re-accredited by NAAC
ISO 9001:2015 Certified Institution
Nanjanapuram, Erode – 638 107.

INFOLINE

EDITORIAL BOARD

CONTENTS

CRYPTOGRAPHIC PROTOCOL ENABLES GREATER COLLABORATION IN DRUG DISCOVERY



MIT researchers have developed a cryptographic system that could help neural networks identify promising drug candidates in massive pharmacological datasets, while keeping the data private. Secure computation done at such a massive scale could enable broad pooling of sensitive pharmacological data for predictive drug discovery. Models can be trained to crunch datasets of known DTIs and then, using that information, find novel drug candidates.

In recent years, pharmaceutical firms, universities and other entities are open to pooling pharmacological data into larger databases that can greatly improve training of these models. Due to intellectual property matters and other privacy concerns, however, these datasets remain limited in scope. Cryptography methods to secure the data are so computationally intensive they don't scale well to datasets beyond, say, tens of thousands of DTIs, which is relatively small. In a paper published today in Science, researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) describe a neural network securely trained and tested on a dataset of more than a million DTIs. The network leverages modern cryptographic tools and optimization techniques to keep the input data private, while running quickly and efficiently at scale.

The team's experiments show the network performs faster and more accurately than existing approaches; it can process massive datasets in days, whereas other cryptographic frameworks would take months. Moreover, the network identified several novel interactions, including one between the leukemia drug imatinib and an enzyme ErbB4 mutations of which have been associated with cancer which could have clinical significance.

People realize that they need to pool their data to greatly accelerate the drug discovery process and enable us, together, to make scientific advances in solving important human diseases, such as cancer or diabetes. But they don't have good ways of doing it, says corresponding author Bonnie Berger, the Simons Professor of Mathematics and a

principal investigator at CSAIL. With this work, we provide a way for these entities to efficiently pool and analyze their data at a very large scale.

"Secret sharing" data

The new paper builds on previous work by the researchers in protecting patient confidentiality in genomic studies, which find links between particular genetic variants and incidence of disease. That genomic data could potentially reveal personal information, so patients can be reluctant to enroll in the studies. In that work, Berger, Cho, and a former Stanford University PhD student developed a protocol based on a cryptography framework called "secret sharing," which securely and efficiently analyzes datasets of a million genomes. In contrast, existing proposals could handle only a few thousand genomes.

Secret sharing is used in multiparty computation, where sensitive data is divided into separate "shares" among multiple servers. Throughout computation, each party will always have only its share of the data, which appears fully random. Collectively, however, the servers can still communicate and perform useful operations on the underlying private data. At the end of the computation, when a result is needed, the parties combine their shares to reveal the result.

A key innovation was reducing the computation needed in training and testing. Existing predictive drug-discovery models represent the chemical and protein structures of DTIs as graphs or matrices. These approaches, however, scale quadratically, or squared, with the number of DTIs in the dataset. Basically, processing these representations becomes extremely computationally intensive as the size of the dataset grows. "While that may be fine for working with the raw data, if you try that in secure computation, it's infeasible," Hie says. The researchers instead trained a neural network that relies on linear calculations, which scale far more efficiently with the data. "We absolutely needed scalability, because we're trying to provide a way to pool data together much larger datasets," Cho says.

The researchers trained a neural network on the STITCH dataset which has 1.5 million DTIs, making it the largest publicly available dataset of its kind. In training, the network encodes each drug compound and protein structure as a simple vector representation. This essentially condenses the complicated structures as 1s and 0s that a computer can easily process. From those vectors, the network then learns the patterns of interactions and non interactions. Fed new pairs of compounds and protein structures, the network then predicts if they'll interact.

The network also has an architecture optimized for efficiency and security. Each layer of a neural network requires some activation function that determines how to send

the information to the next layer. In their network, the researchers used an efficient activation function called a rectified linear unit (ReLU). This function requires only a single, secure numerical comparison of an interaction to determine whether to send (1) or not send (0) the data to the next layer, while also never revealing anything about the actual data. This operation can be more efficient in secure computation compared to more complex functions, so it reduces computational burden while ensuring data privacy.

"The reason that's important is we want to do this within the secret sharing framework … and we don't want to ramp up the computational overhead," Berger says. In the end, "no parameters of the model are revealed and all input data the drugs, targets, and interactions are kept private."

Finding interactions

The researchers pitted their network against several state-of-the-art, plaintext (unencrypted) models on a portion of known DTIs from DrugBank, a popular dataset containing about 2,000 DTIs. In addition to keeping the data private, the researchers' network outperformed all of the models in prediction accuracy. Only two baseline models could reasonably scale to the STITCH dataset, and the researchers' model achieved nearly double the accuracy of those models.

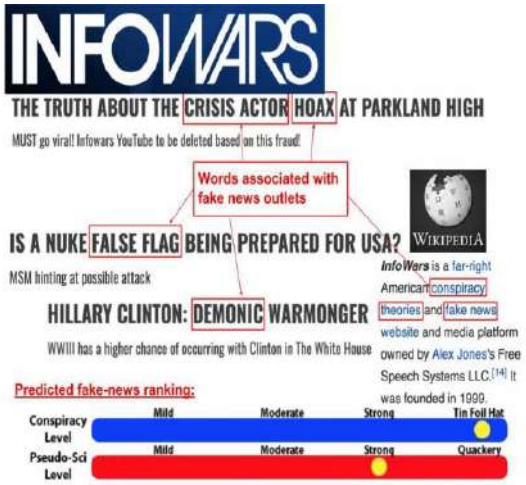
The researchers also tested drug-target pairs with no listed interactions in STITCH, and found several clinically established drug interactions that weren't listed in the database but should be. In the paper, the researchers list out the top strongest predictions including droloxifene and an estrogen receptor, which reached phase III clinical trials as a treatment for breast cancer; and seocalcitol and a vitamin D receptor to treat other cancers. Cho and Hie independently validated the highest-scoring novel interactions via contract research organizations.

The work could be "revolutionizing" for predictive drug discovery, says Artemis Hatzigeorgiou, a professor of bioinformatics at the University of Thessaly in Greece. "Having entered the era of big data in pharmacogenetics, it is possible for the first time to retrieve a dataset of this unprecedented big size from patient data. Similar to the learning procedure of a human brain, artificial neural networks need a critical mass of data in order to provide confident decisions," Hatzigeorgiou says. "Now is possible the use of millions of data to train an artificial neural network toward the identification of unknown drug-target interactions. Under such conditions, it is not a surprise that this trained model outperforms all existing methods on drug discovery."

POORANI G
III B.Sc. (Computer Technology)

DETECTING FAKE NEWS AT ITS SOURCE

Lately the fact-checking world has been in a bit of a crisis. Sites like Politifact and Snopes have traditionally focused on specific claims, which is admirable but tedious by the time they've gotten through verifying or debunking a fact, there's a good chance it's already travelled across the globe and back again.



Social media companies have also had mixed results limiting the spread of propaganda and misinformation. Facebook plans to have 20,000 human moderators by the end of the year, and is putting significant resources into developing its own fake-news-detecting algorithms. Researchers from MIT's Computer Science and Artificial Intelligence Lab (CSAIL) and the Qatar Computing Research Institute (QCRI) believe that the best approach is to focus not only on individual claims, but on the news sources themselves. Using this tack, they've demonstrated a new system that uses machine learning to determine if a source is accurate or politically biased.

"If a website has published fake news before, there's a good chance they'll do it again," says postdoc Ramy Baly, the lead author on a new paper about the system. "By automatically scraping data about these sites, the hope is that our system can help figure out which ones are likely to do it in the first place." Baly says the system needs only about 150 articles to reliably detect if a news source can be trusted meaning that an approach like theirs could be used to help stamp out new fake-news outlets before the stories spread too widely.

The system is a collaboration between computer scientists at MIT CSAIL and QCRI, which is part of the Hamad Bin Khalifa University in Qatar. Researchers first took data from Media Bias/Fact Check (MBFC), a website with human fact-checkers who analyze the accuracy and biases of more than 2,000 news sites; from MSNBC and Fox News; and from low-traffic content farms. They then fed those data to a machine learning algorithm, and programmed it to classify news sites the same way as MBFC. When given a new news outlet, the system was then 65 percent accurate at detecting whether it has a high, low or medium level of factuality, and roughly 70 percent accurate at detecting if it is left-leaning, right-leaning, or moderate.

The team determined that the most reliable ways to detect both fake news and biased reporting were to look at the common linguistic features across the source's stories, including sentiment, complexity, and structure. For example, fake-news outlets were found to be more likely to use language that is hyperbolic, subjective, and emotional. In terms of bias, left-leaning outlets were more likely to have language that related to concepts of harm/care and fairness/reciprocity, compared to other qualities such as loyalty, authority, and sanctity. Co-author Preslav Nakov, a senior scientist at QCRI, says that the system also found correlations with an outlet's Wikipedia page, which it assessed for general longer is more credible as well as target words such as "extreme" or "conspiracy theory." It even found correlations with the text structure of a source's URLs: Those that had lots of special characters and complicated subdirectories, for example, were associated with less reliable sources.

"Since it is much easier to obtain ground truth on sources [than on articles], this method is able to provide direct and accurate predictions regarding the type of content distributed by these sources," says Sibel Adali, a professor of computer science at Rensselaer Polytechnic Institute who was not involved in the project. Nakov is quick to caution that the system is still at work in progress, and that, even with improvements in accuracy, it would work best in conjunction with traditional fact-checkers.

"If outlets report differently on a particular topic, a site like Politifact could instantly look at our fake news scores for those outlets to determine how much validity to give to different perspectives," says Nakov. Baly and Nakov co-wrote the new paper with MIT Senior Research Scientist James Glass alongside graduate students Dimitar Alexandrov and Georgi Karadzhov of Sofia University. The team will present the work later this month at the 2018 Empirical Methods in Natural Language Processing (EMNLP) conference in Brussels, Belgium.

The researchers also created a new open-source dataset of more than 1,000 news sources, annotated with factuality and bias scores, that is the world's largest database of its kind. As next steps, the team will be exploring whether the English-trained system can be adapted to other languages, as well as to go beyond the traditional left/right bias to explore region-specific biases "This direction of research can shed light on what untrustworthy websites look like and the kind of content they tend to share, which would be very useful for both web designers and the wider public," says Andreas Vlachos, a senior lecturer at the University of Cambridge who was not involved in the project.

Nakov says that QCRI also has plans to roll out an app that helps users step out of their political bubbles, responding to specific news items by offering users a collection of articles that span the political spectrum. “It’s interesting to think about new ways to present the news to people,” says Nakov. “Tools like this could help people give a bit more thought to issues and explore other perspectives that they might not have otherwise considered."

G.LOKESHWAR
III B.Sc. (Information Technology)

A STEP TOWARD PERSONALIZED, AUTOMATED SMART HOMES

Developing automated systems that track occupants and self-adapt to their preferences is a major next step for the future of smart homes. When you walk into a room, for instance, a system could set to your preferred temperature. Or when you sit on the couch, a system could instantly flick the television to your favourite channel. But enabling a home system to recognize occupants as they move around the house is a more complex problem. Recently, systems have been built that localize humans by measuring the reflections of wireless signals off their bodies. But these systems can’t identify the individuals. Other systems can identify people, but only if they’re always carrying their mobile devices. Both systems also rely on tracking signals that could be weak or get blocked by various structures.

MIT researchers have built a system that takes a step towards a fully automated smart home by identifying occupants, even when they’re not carrying mobile devices. The system, called Duet, uses reflected wireless signals to localize individuals. But it also incorporates algorithms that ping nearby mobile devices to predict the individuals’ identities, based on who last used the device and their predicted movement trajectory. It also uses logic to figure out who’s who, even in signal-denied areas.

“Smart homes are still based on explicit input from apps or telling Alexa to do something. Ideally, we want homes to be more reactive to what we do, to adapt to us,” says Deepak Vasisht, a PhD student in MIT’s Computer Science and Artificial Intelligence Laboratory (CSAIL) and lead author on a paper

describing the system that was presented at last week's Ubicomp conference. "If you enable location awareness and identification awareness for smart homes, you could do this automatically. Your home knows it's you walking, and where you're walking, and it can update itself." Experiments done in a two-bedroom apartment with four people and an office with nine people, over two weeks, showed the system can identify individuals with 96 percent and 94 percent accuracy, respectively, including when people weren't carrying their smartphones or were in blocked areas.

But the system isn't just novelty. Duet could potentially be used to recognize intruders or ensure visitors don't enter private areas of your home. Moreover, Vasisht says, the system could capture behavioral-analytics insights for health care applications. Someone suffering from depression, for instance, may move around more or less, depending on how they're feeling on any given day. Such information, collected over time, could be valuable for monitoring and treatment. "In behavioural studies, you care about how people are moving over time and how people are behaving," Vasisht says. "All those questions can be answered by getting information on people's locations and how they're moving."

The researchers envision that their system would be used with explicit consent from anyone who would be identified and tracked with Duet. If needed, they could also develop an app for users to grant or revoke Duet's access to their location information at any time, Vasisht adds. Co-authors on the paper are: Dina Katabi, the Andrew and Erna Viterbi Professor of Electrical Engineering and Computer Science; former CSAIL researcher Anubhav Jain '16; and CSAIL PhD students Chen-Yu Hsu and Zachary Kabelac.

Tracking and identification

Duet is a wireless sensor installed on a wall that's about a foot and a half squared. It incorporates a floor map with annotated areas, such as the bedroom, kitchen, bed, and living room couch. It also collects identification tags from the occupants' phones. The system builds upon a device-based localization system built by Vasisht, Katabi, and other researchers that tracks individuals within tens of centimetres, based on wireless signal reflections from their devices. It does so by using a central node to calculate the time it takes the signals to hit a person's device and travel back. In experiments, the system was able to pinpoint where people were in a two-bedroom apartment and in a café.

The system, however, relied on people carrying mobile devices. "But in building [Duet] we realized, at home you don't always carry your phone," Vasisht says. "Most people leave devices on desks or tables, and walk around the house." The researchers combined

their device-based localization with a device-free tracking system, called WiTrack, developed by Katabi and other CSAIL researchers, that localizes people by measuring the reflections of wireless signals off their bodies.

Duet locates a smartphone and correlates its movement with individual movement captured by the device-free localization. If both are moving in tightly correlated trajectories, the system pairs the device with the individual and therefore, knows the identity of the individual. To ensure Duet knows someone's identity when they're away from their device, the researchers designed the system to capture the power profile of the signal received from the phone when it's used. That profile changes, depending on the orientation of the signal, and that change be mapped to an individual's trajectory to identify them. For example, when a phone is used and then put down, the system will capture the initial power profile. Then it will estimate how the power profile would look if it were still being carried along a path by a nearby moving individual. The closer the changing power profile correlates to the moving individual's path, the more likely it is that individual owns the phone.

Logical thinking

One final issue is that structures such as bathroom tiles, television screens, mirrors, and various metal equipment can block signals. To compensate for that, the researchers incorporated probabilistic algorithms to apply logical reasoning to localization. To do so, they designed the system to recognize entrance and exit boundaries of specific spaces in the home, such as doors to each room, the bedside, and the side of a couch. At any moment, the system will recognize the most likely identity for each individual in each boundary. It then infers who is who by process of elimination.

Suppose an apartment has two occupants: Alisha and Betsy. Duet sees Alisha and Betsy walk into the living room, by pairing their smart phone motion with their movement trajectories. Both then leave their phones on a nearby coffee table to charge Betsy goes into the bedroom to sleep, Alisha stays on the couch to watch television. Duet infers that Betsy has entered the bed boundary and didn't exit, so must be on the bed. After a while, Alisha and Betsy move into the kitchen and the signal drops. Duet reasons that two people are in the kitchen but it doesn't know their identities. When Betsy returns to the living room and picks up her phone, however, the system automatically re-tags the individual as Betsy. By process of elimination, the other person still in the kitchen is Alisha. "There are blind spots in homes where systems won't work. But, because you have logical framework, you can make these inferences," Vasisht says.

"Duet takes a smart approach of combining the location of different devices and associating it to humans, and leverages device-free localization techniques for localizing humans," says Ranveer Chandra, a principal researcher at Microsoft, who was not involved in the work. "Accurately determining the location of all residents in a home has the potential to significantly enhance the in-home experience of users. The home assistant can personalize the responses based on who all are around it; the temperature can be automatically controlled based on personal preferences, thereby resulting in energy savings. Future robots in the home could be more intelligent if they knew who was where in the house. The potential is endless."

KEERTHANA R
III B.Sc. (Computer Technology)

ADVANCED DATA ANALYTICS

Organizations and companies have been using basic data analytics for years to uncover simple insights and trends. The appetite for more data and better analytics has grown over the years, and now most modern organizations track and record nearly all types of data: transactional, clickstream, social media, audio, video, sensor, text, image, and so on. This ever-increasing volume of data, along with the diversity of data sources, makes the process of extracting useful information and insights an ever more challenging and complex endeavour.

To meet this challenge, organizations and companies have turned to advanced data analytics as an overarching approach to finding the value hidden in the mountains of data that they are rapidly accumulating. Gartner defines advanced data analytics as autonomous or semi-autonomous data and content examination using sophisticated quantitative and qualitative techniques and tools with the goal of discovering deep insights and subtle patterns

and of making predictions and recommendations. These techniques tend to be interdisciplinary and span fields such as

- Data Mining
- Machine Learning
- Pattern Matching
- Visualization
- Semantic Analysis
- Sentiment Analysis
- Networking
- Multivariate Statistics
- Graph Analysis
- Event Processing
- Neural Networks.

Most standard learning algorithms assume or expect datasets to have balanced class distributions or equal misclassification costs. However, in "Learning from Imbalanced Data," Edwardo A. Garcia and Haibo He argues that standard learning algorithms fail to properly represent the distributive characteristics of datasets in some fields (for example, biomedicine) that exhibit unequal distribution between classes. This article aims to provide a survey of the current research developments about the imbalanced learning problem and review state-of-the-art solutions. It also highlights opportunities and challenges for learning from imbalanced data.

Extracting valuable information from petabytes of data requires new clustering algorithms that are scalable, less computationally intensive, and ready for implementation on optimized, large-scale interactive computational frameworks. In "Fuzzy Based Scalable Clustering Algorithms for Handling Big Data Using Apache Spark," Neha Bharill, Aruna Tiwari, and Aayushi Malviya develop an algorithm for implementation on an Apache Spark Cluster to address the challenges associated with big data clustering. The authors point out that their work achieves a significant reduction in runtime for the clustering of huge amounts of data, without compromising the quality of clustering results. Optimization techniques eliminate the need for storing large membership data matrices during the execution of the proposed algorithm, resulting in shorter runtime.

"Massive Social Network Analysis: Mining Twitter for Social Good" analyzes Twitter's vast quantity of unstructured data. David Ediger and his colleagues present GraphCT, a Graph Characterization Toolkit for massive graphs representing social network data. GraphCT analyzed graphs representing Twitter's public crisis data stream and revealed interesting characteristics of Twitter users' interactions. This allows for identifying influential sources and ranking conversations, thus enabling analysts to focus on a manageable number of conversations.

Healthcare policy is one of the most prominent data analytics application areas. In

"Improving Healthcare with Interactive Visualization," Bradford W. Hesse, Ben Shneiderman, and Catherine Plaisant map the healthcare information into three domains personal health, clinical health, and public health and highlight the central role of information visualization and visual analytics in enabling patients, clinicians, and public health policymakers to make better decisions. The article outlines seven practical challenges in the three health domains and highlights the opportunities for information visualization tools, techniques, and analytics that could help mitigate these challenges and lead to improved healthcare.

Alexander Brodsky and his colleagues consider smart manufacturing a data analytics application area in "Analysis and Optimization in Smart Manufacturing based on a Reusable Knowledge Base for Process Performance Models." They posit that to analyze the performance of complex production lines (such as car production lines), different types of analysis and optimization capabilities such as descriptive, diagnostic, predictive, and prescriptive analytics are needed. Each of these capabilities is based on a variety of data that is filtered and aggregated over time and space; for example, descriptive analytics uses temporal sensor data that include line speeds, CO_2 emissions, and water consumption. The article proposes an architectural design and framework for fast development of software solutions for descriptive, diagnostic, predictive, and prescriptive analytics of dynamic production processes.

As the world moves rapidly into the digital age, individuals, organizations, and companies are being inundated with data. Advanced data analytics offers a plethora of opportunities for researchers, policy analysts, and business managers to innovate and develop tools, techniques, strategies, policies, and software products to extract valuable insights from data. We hope this Computing Now theme issue inspires more research in this rich field.

VARSHA R

II B.Sc. (Information Technology)

MAKING 'SMART HEADLIGHTS' WITH MACHINE LEARNING

For example, the headlight could reduce the amount of light pointed at an oncoming car while simultaneously increasing the illumination of an upcoming road sign. Or it

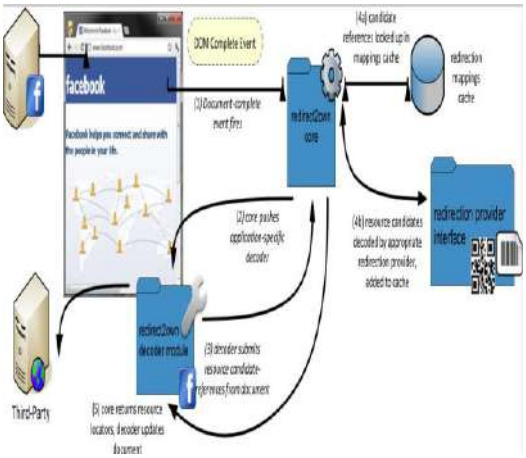
could detect nearby pedestrians and alert the driver by highlighting their body while avoiding shining light directly into their eyes.

The challenge to making such a headlight isn't necessarily creating the different patterns of light. It's teaching the car how to automatically recognize the surrounding environment and make the patterns on its own. It's a problem that HASCO Vision Technology the largest automotive lamp manufacturer in China is turning to Li to help them solve with machine learning.

Many car companies are using cameras and machine learning to help control their own versions of self-driving cars, so Li is far from alone in this regard. Machine learning algorithms, however, need vast amounts of data to learn from, and many of the datasets and algorithms already created for this purpose have focused on daytime driving.

IYSWARIYA R
II B.Sc. (Information Technology)

A NEW APPROACH TO PROTECT THE INTELLECTUAL PROPERTY OF USER-UPLOADED CONTENT

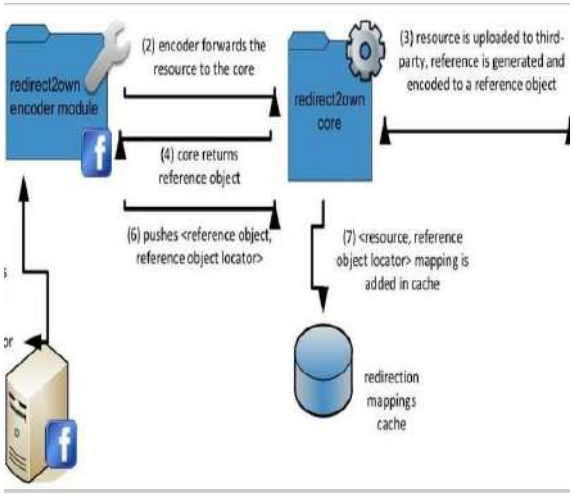


Every day, millions of users upload content on social media media platforms, including text, pictures and videos. While the creators of this content typically retain intellectual property (IP) rights once it is shared online, they often grant services hosting the content a broad license to use it, possibly in undesirable ways. For example, Facebook claims a transferable, sub-licensable and royalty-free license on all user-provided content, which is applicable worldwide. This particularly affects photographers and other creative professionals whose work is shared online, taking away their right to decide where it will be featured.

Researchers at Columbia University, Georgia Institute of Technology and Stevens Institute of Technology have devised a new method to safeguard intellectual property (IP) rights of content uploaded online. Their approach was recently presented in a paper pre-

published on arXiv. "The research was motivated by the fact that users frequently need to agree to terms of service (TOS), which grant excessive rights to services, before being able to share pictures through them,"

Georgios Portokalidis, one of the researchers who carried out the study told Tech Xplore, "Our goal was to investigate new techniques that would allow users to share their photos through such services, while actually hosting their pictures in services with more favorable terms or on their own servers."



The approach proposed by Portokalidis and his colleagues builds on technologies such as QR codes which can effectively encode information in images, as well as browser extensions that can transparently modify website content. Essentially, their design works by decoupling user data from social networking services, without incurring any loss of functionality to the user.

"Our approach is based on uploading QR codes instead of the original pictures on the social networking services," Portokalidis said. "These QR codes contain the encoded location of the actual picture. A browser extension transparently identifies these QR codes and replaces them with the original picture, as users browse them on social networking sites. The same extension can be used to transparently intercept uploads to a service (e.g., Facebook), uploading the pictures instead to a more favourable second service (e.g., Flickr). The first service gets QR code images instead."

S.AISWARYA

II B.Sc. (Computer Technology)

3-D METAL PRINTING

While 3-D printing has been around for decades, it has remained largely in the domain

of hobbyists and designers producing one-off prototypes. And printing objects with anything other than plastics in particular, metal has been expensive and painfully slow. Now, however, it's becoming cheap and easy enough to be a potentially practical way of manufacturing parts. If widely adopted, it could change the way we mass-produce many products.

- Breakthrough New printers can make metal objects quickly and cheaply.
- Why It Matters The ability to make large and complex metal objects on demand could transform manufacturing.
- Key Players Markforged, Desktop Metal, GE

In the short term, manufacturers wouldn't need to maintain large inventories. They could simply print an object such as a replacement part for an aging car, whenever someone needs it. In the longer term, large factories that mass-produce a limited range of parts might be replaced by smaller ones that make a wider variety, adapting to customers' changing needs.

The technology can create lighter, stronger parts, and complex shapes that aren't possible with conventional metal fabrication methods. It can also provide more precise control of the microstructure of metals. In 2017, researchers from the Lawrence Livermore National Laboratory announced they had developed a 3-D-printing method for creating stainless-steel parts twice as strong as traditionally made ones.

Another Boston-area startup, Desktop Metal, began to ship its first metal prototyping machines in December 2017. It plans to begin selling larger machines, designed for manufacturing, that are 100 times faster than older metal printing methods. The printing of metal parts is also getting easier. Desktop Metal now offers software that generates designs ready for 3-D printing. Users tell the program the specs of the object they want to print, and the software produces a computer model suitable for printing.

GE which has long been a proponent of using 3-D printing in its aviation products has a test version of its new metal printer that is fast enough to make large parts. The company plans to begin selling the printer in 2018.

SRIDHARSHINI R K
III B.Sc. (Information Technology)

BIG DATA TECHNOLOGIES

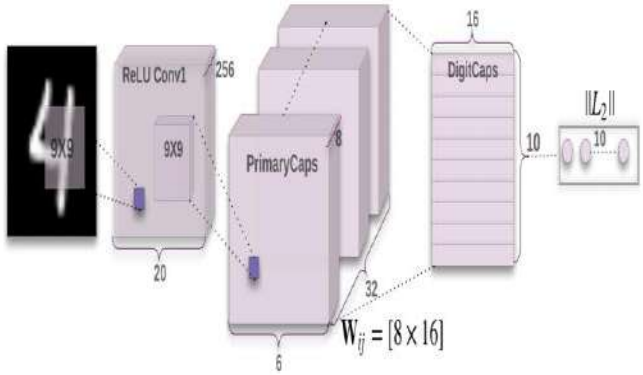
1. **Predictive analytics**: software and/or hardware solutions that allow firms to discover, evaluate, optimize, and deploy predictive models by analyzing big data sources to improve business performance or mitigate risk.
2. **NoSQL databases**: key-value, document, and graph databases.

3. **Search and knowledge discovery**: tools and technologies to support self-service extraction of information and new insights from large repositories of unstructured and structured data that resides in multiple sources such as file systems, databases, streams, APIs, and other platforms and applications.
4. **Stream analytics**: software that can filter, aggregate, enrich, and analyze a high throughput of data from multiple disparate live data sources and in any data format.
5. **In-memory data fabric**: provides low-latency access and processing of large quantities of data by distributing data across the dynamic random access memory (DRAM), Flash, or SSD of a distributed computer system.
6. **Distributed file stores**: a computer network where data is stored on more than one node, often in a replicated fashion, for redundancy and performance.
7. **Data virtualization**: a technology that delivers information from various data sources, including big data sources such as Hadoop and distributed data stores in real-time and near-real time.
8. **Data integration**: tools for data orchestration across solutions such as Amazon Elastic MapReduce (EMR), Apache Hive, Apache Pig, Apache Spark, MapReduce, Couchbase, Hadoop, and MongoDB.
9. **Data preparation**: software that eases the burden of sourcing, shaping, cleansing, and sharing diverse and messy data sets to accelerate data's usefulness for analytics.
10. **Data quality**: products that conduct data cleansing and enrichment on large, high-velocity data sets, using parallel operations on distributed data stores and databases.

J.JANSI RANI
III B.Sc. (Computer Technology)

CAPSULE NETWORK

Every few days there is an advancement in the field of Neural Networks. Some brilliant minds are working on this field. You can pretty much assume every paper on this topic is almost ground breaking or path changing. MNIST is a database of tagged handwritten digit images. Results are showing a significant increase in the performance in case of overlapped digits. The capsules are particularly good at handling different types of visual stimulus and encoding things like pose (position, size, orientation), deformation, velocity, albedo, hue, texture etc. The brain must have a mechanism for "routing" low level visual information to what it believes is the best capsule for handling it.



Capsule is a nested set of neural layers. So in a regular neural network you keep on adding more layers. In CapsNet you would add more layers inside a single layer. Or in other words nest a neural layer inside another. The state of the neurons inside a capsule capture the above properties of one entity inside an image. Capsule outputs a vector to represent the existence of the entity. The orientation of the vector represents the properties of the entity. The vector is sent to all possible parents in the neural network. For each possible parent a capsule can find a prediction vector. Prediction vector is calculated based on multiplying it's own weight and a weight matrix. Whichever parent has the largest scalar prediction vector product, increases the capsule bond. Rest of the parents decrease their bond. This routing by agreement method is superior than the current mechanism like max-pooling. Max pooling routes based on the strongest feature detected in the lower layer. Apart from dynamic routing, CapsNet talks about adding squashing to a capsule. Squashing is a non-linearity. So instead of adding squashing to each layer like how you do in CNN, you add the squashing to a nested set of layers. So the squashing function gets applied to the vector output of each capsule.

S.AISWARYA

II B.Sc. (Computer Technology)

A GUIDE TO PROTECTING YOUR PRIVACY ON WHATSAPP

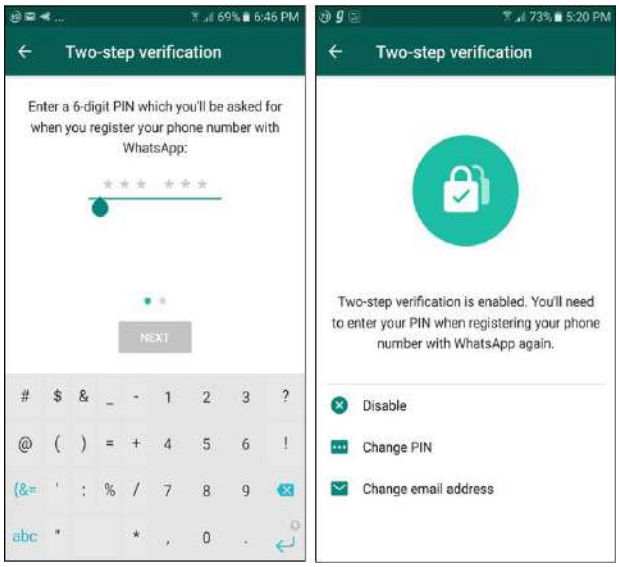
Whatsapp is a popular messaging service with over a billion users. With the introduction of end-to-end encryption in 2016, WhatsApp has made sure that message details do not fall into the wrong hands. However, there are more steps that you can take to make your messages more secure. Here are 7 essential tips to keep in mind:

Manage WhatsApp Web effectively

While the web version of WhatsApp has made reading and replying to messages easier and it also lets you access your messages even if your android is not in near sight. All

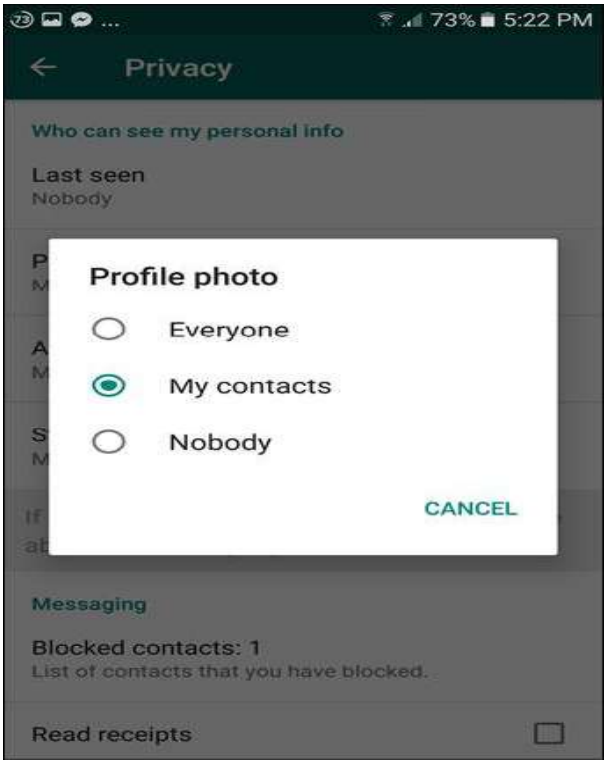
that you need to do is log in once using the QR code. While this is convenient if you are using your own personal laptop or desktop to access Whatsapp, on a public computer, be sure to log out from all computers.

Use two-factor verification



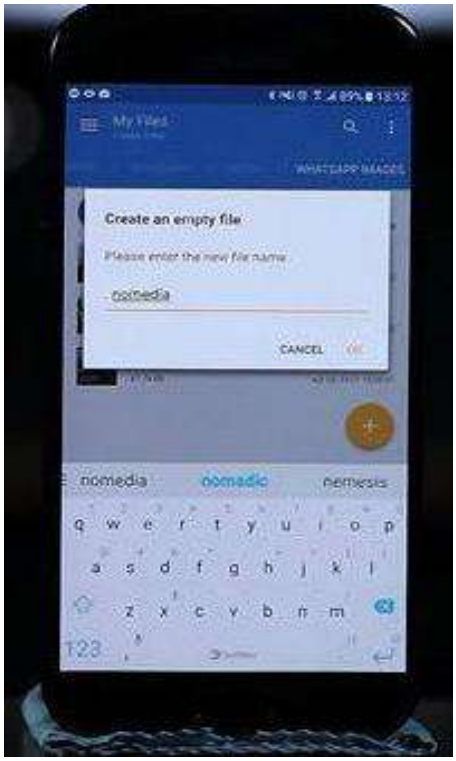
WhatsApp has adequate security but only if using the two-factor authentication. SIM cloning is rather easy these days, so this method ensures that nobody else can set up your WhatsApp on their phone without knowing the secret 6-digit code. Settings can be accessed through Account:Two-Step Verification. Once enabled, any re-verification must be done with this number. As an additional measure, you can also set up an alternate email address which will send a link to your email once this authentication is disabled.

Restrict access to your profile picture



What made WhatsApp popular was the fact that anybody can be texted with just a contact number. Yet one thing you need to be aware of is that your profile picture can land in the wrong hands. But WhatsApp has a function that allows you to restrict who sees your profile picture. All you have to do is head over to Privacy and switch from the default Everybody to My Contacts. With this function, only contacts which you have saved will be able to see your profile picture. You can also opt for Nobody, should you wish to keep your profile picture completely hidden.

Hide photos from gallery



WhatsApp saves all media items sent to your phone automatically to your gallery. But, it is possible for WhatsApp media to be easily hidden from the gallery, keeping whatever was sent to you a secret. All you need to do is go to the WhatsApp image folder using the File Explorer and create a file called nomedia. Once you've rebooted your phone, all your WhatsApp media will be hidden. Alternatively, you can also download an app named Hide image, video of WhatsApp which hides the required folders.

Lock WhatsApp

You also have the ability to lock the app from prying eyes using a third party app. If your messages contain highly sensitive information than AppLock. You can choose to have a lock on with a PIN or a pattern. The app also lets you lock individual apps and various system switches such as Bluetooth, Wi-Fi and so on.

Deactivate account on a lost or stolen phone

If your phone gets lost or stolen, the WhatsApp account can also be deactivated along with the SIM deactivation. However, remote WhatsApp deactivation is not as yet supposed meaning that deactivating it is a bit

long. You would need to send an e-mail to WhatsApp support with the phrase LOST/STOLEN: PLEASE DEACTIVATE MY ACCOUNT (phone number) as the subject line. Be sure that your number is in the correct international format.

Be aware of scams

The above points ensure that your phone apps remain secure. However, you should always take care with any suspicious links. As contents are not actively monitored, WhatsApp is fast turning into a den of suspicious links and misleading content.

SELVA BHARATHI A
II B.Sc. (Computer Technology)

GOOGLE ASSISTANT MAKES A RESERVATION BY PHONE

The Google Assistant has gone from a few tricks pony to full-on sci-fi-esque post Google's I/O 2018 developer conference. Google is adding so many features to its personal voice assistant, it wouldn't be wrong to say that the Google Assistant is en route to get its biggest ever revamp, this year. Many of the features that Google announced at I/O 2018, for the Google Assistant, aren't ready for prime time yet and would be added in due course of time.

Being conversational is one thing (even Amazon's Alexa can do that) but Google's new voices for the Google Assistant are dangerously close to mimicking the real thing. They sound dangerously natural. What with Google's massive scale vis-a-vis AI and machine learning, it was only about time that we started seeing such things come to life. So bottom-line is the Google Assistant is getting six new voices, in addition to its poster child Holly, and R&B singer and songwriter John

Legend will be coming to it by the end of this year, Google has confirmed.

If you're an Android user, you can start using the brand new Google Assistant, right now. The update is now rolling out to all Android devices that support the Google Assistant, including phone, tablet, and Google Home hardware. Interestingly, the six new voices aren't part of the Android P Beta program in the case of the Pixel phones for now. I am not sure whether that's a global thing or something that's specific to region. But, here's the thing, you can't get the Assistant's new voices if you're on a Pixel that's running Android P Beta. At least, it didn't work for me. If you were able to get it, please make sure to leave a comment down below.

Every other Android device, that supports the Google Assistant, however can help you choose your new voices. And it's a cross-platform and cross-device thing, which means if you change the voice of your Assistant on one platform/device, the change will reflect across all your platforms/devices. Even the Pixel that's running Android P Beta. For now, the only way that you can get different Google Assistant voices, is you choose to tie your devices with different Google accounts.

Now that we have all of that out of our way, here's how you can get access to all of the Google Assistant's new voices:

1. Open the Google Assistant app by holding down the home button on your phone.

2. Click the blue icon at the top right of the Google Assistant box. This will take you to the explore page. Click the three dots icon in the top right-hand corner to bring out the menu. In the menu, choose Settings. This will take you to the Assistant settings page.

3. In the Assistant settings page, hit the Preferences option.

4. Your Preferences column should have an Assistant voice tab (if not, you'll need to update your Google app to the latest version).

5. Hit the Assistant voice tab and you'll be taken to the new Google Assistant voice options.

6. You will now have eight voice options to choose from total. Google doesn't have a name for each of its new voices yet. Neither does it tell you anything about what to expect from them via hints. You'll have to manually tap on each of these voices to hear them out and then you can choose to set one as default.

Adding new voices isn't the only thing that Google announced at I/O. The most striking announcement was Duplex, a smart feature that lets the Assistant make phone calls for you on your behalf. Duplex can "make

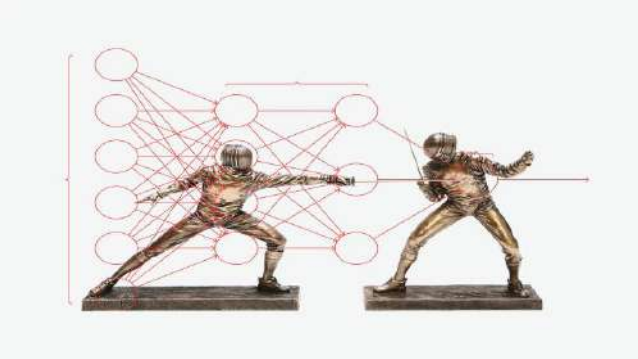
reservations, schedule appointments, and get holiday hours from businesses. Just provide the date and time, and your Assistant will call the business to coordinate for you." Just to ensure people aren't uncomfortable the feature is undoubtedly creepy Google has confirmed that Duplex will come with disclosure meaning it will let the other side know that they're talking to a bot possibly before starting the conversation.

In addition to new voices and Duplex, the Google Assistant will support a new feature that Google is calling Continued Conversation. This will allow the Assistant to hold more natural conversations with you. Also, you won't need to say hey to Google after every query. You can continue to chat with the Assistant without saying the wake word, and a simple thank you, would be a cue for it to stop. Also, just to ensure, kids aren't acting too bossy around it, Google is pushing for a feature called (pretty) please for obvious reasons.

MOHANRAJ E

III B.Sc. (Information Technology)

DUELING NEURAL NETWORKS



Artificial intelligence is getting very good at identifying things: show it a million pictures, and it can tell you with uncanny accuracy which ones depict a pedestrian crossing a street. But AI is hopeless at generating images of pedestrians by itself. If it could do that, it would be able to create gobs of realistic but synthetic pictures depicting pedestrians in various settings, which a self-driving car could use to train itself without ever going out on the road. The problem is, creating something entirely new requires imagination and until now that has perplexed AIs.

The solution first occurred to Ian Goodfellow, then a PhD student at the University of Montreal, during an academic argument in a bar in 2014. The approach, known as a generative adversarial network, or GAN, takes two neural networks the simplified mathematical models of the human brain that underpin most modern machine learning and pits them against each other in a digital cat-and-mouse game.

Both networks are trained on the same data set. One, known as the generator, is tasked with creating variations on images it's already seen perhaps a picture of a pedestrian with an extra arm. The second, known as the discriminator, is asked to identify whether the example it sees is like the images it has been trained on or a fake produced by the generator basically, is that three-armed person likely to be real?

Over time, the generator can become so good at producing images that the discriminator can't spot fakes. Essentially, the generator has been taught to recognize, and then create, realistic-looking images of pedestrians. The technology has become one of the most promising advances in AI in the past decade, able to help machines produce results that fool even humans.

GANs have been put to use creating realistic-sounding speech and photorealistic fake imagery. In one compelling example, researchers from chipmaker Nvidia primed a GAN with celebrity photographs to create hundreds of credible faces of people who don't exist. Another research group made not-unconvincing fake paintings that look like the works of van Gogh. Pushed further, GANs can reimagine images in different ways making a sunny road appear snowy or turning horses into zebras.

The results aren't always perfect: GANs can conjure up bicycles with two sets of handlebars, say, or faces with eyebrows in the wrong place. But because the images and sounds are often startlingly realistic, some experts believe there's a sense in which GANs are beginning to understand the underlying structure of the world they see and hear. And that means AI may gain, along with a sense of imagination, a more independent ability to make sense of what it sees in the world.

SURESHKUMAR
I B.Sc. (Computer Technology)

---

iTwin

USB flash drive is a data storage device that consists of flash memory with an integrated Universal Serial Bus (USB) interface. A Cloud Storage is also a similar case where in the data is stored remotely and is accessed whenever needed.

The drawback of this USB flash drives is small size cause of which it can easily be misplaced or lost. This is a particular problem if the data it contains is sensitive and confidential. In Cloud Storages the data can be stolen or misused if the username and password of an account to access the storage is hacked by someone. There are many more drawbacks like limited storage, no security, back up, temp files, no remote disable etc.

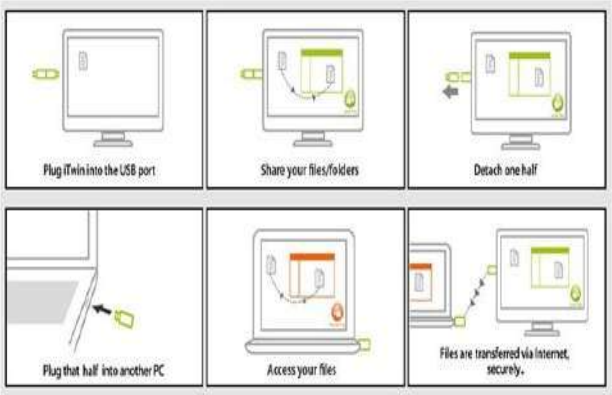
iTwin is a 'limitless' secure USB device that enables users to access, edit & share all their files & media between any two online computers anywhere in the world. The only limit is the size of your computer's hard drive. iTwin is an innovative solution that allows remote file access without the security and privacy risks of the cloud and USB flash drives. It's very easy to access as a USB device and no special installation is required. iTwin uses thoroughly analyzed crypto protocols and algorithms(AES 256 bit encryption). It has features like bi-directional file access, no temp files, remote disable, smart key generation, password support and twintrust authentication

What is an iTwin?

iTwin is a revolutionary new file sharing and remote access device brought to you by a company called iTwin. It's like two ends of a cable, without the cable. It's as simple to use as a flash drive. It's literally plug and play. iTwin can connect any two online computers anywhere in the world. iTwin enables you to have access to any or all of your home computer's files and folders while you're on-the-go. Similarly, you can also use iTwin to access to any or all of your office computer's files and folders while on-the-go. There's no in-built limit to the amount of storage you can access with iTwin while you're on-the-go. The only limit is the size of your hard drives. The only other limit is the speed of your Internet connection. The faster it is, the better your experience. You can select files for accessing later on-the-go, or you can edit them remotely, without the files leaving your computer. You can also back-up files to your home or office computer while you're out on-the-go. It's so easy, it's unbelievable.

Who invented iTwin?

Engineering from IIT in Chennai and a Masters degree from IISc in Bangalore, Lux worked first as an IT security researcher at the Institute of Systems Science, Singapore and then as senior researcher at Kent Ridge Digital Labs and the Institute for Infocomm Research. Lux specializes in PKI implementations, efficient digital certificate revocations and usable security. Lux was pursuing a part-time MBA at NUS Business School, Singapore, but put studies on hold because of potential of iTwin



How to use iTwin?

When you connect iTwin, to see a regular window pop-up, just as you would if you plugged in a regular USB flash drive. Drag and drop files and folders into this window to share them - as many as you want. Leave your computer with one half of iTwin connected to it. Detach the other half of iTwin and take it with you. Wherever you go, you can remotely access the shared files, simply by plugging the half you are carrying into any online Windows computer, anywhere. iTwin allows you to transfer files to or from your home computer. Or your office computer. Or your friend's, or your colleague's iTwin also allows to edit the shared files on a remote computer, while keeping them on that remote computer.

Unique Features

1. Smart key generation

Two iTwins together generate a random 256-bit AES key, everytime they are physically paired and plugged into a computer. Smart Key generation is assisted by the computer to add randomness. Smart crypto key resides only on the two halves of the paired iTwin. Smart Crypto key is used for encrypting all data traffic between two iTwins.

2. No "Temp Files"

Unplug iTwin and all temp files are purged automatically. This is especially useful when using a computer that does not belong to you.

3. Password support

iTwin even provides password support. Passwords of any length can be set. Unlike other web services, iTwin's password is stored on itself, not on any server. If you forget your password, simply pair both halves of your iTwin, plug them into a computer and set a new password. No need for tech support.

4. Bi-Directional File Access

When using iTwin, the connection between the 2 computers is completely symmetrical. Access, copy, backup & remotely edit files on computer A from computer B, and on B from A. You will have access to files on both A and B.

RIDDLES

1. I am a word of 5 letters!
   Everybody eats me!
   If you remove my first letter, I will be a form of energy!!!
   If you remove my first 2 letters, I will be needed for living!!!
   If you remove my first 3 letter, I will be a preposition and near you!!!
   If you remove my 4 first letters, I will be a drink for you!!!

Answer : WHEAT

I am a word of 5 letters -WHEAT

Everybody eats me

If you remove my first letter, I will be a form of energy- HEAT

If you remove my first 2 letters, I will be needed for living-EAT

If you remove my first 3 letter, I will be a preposition and near you-AT

If you remove my 4 first letters, I will be a drink for you-T(tea)

2. Identify the names of various bikes given in image below :

Answers:

1. Fiero	11. Star City
2. Scooty	12. Shine
3. Fazer	13. Rajdoot
4. Discover	14. 390
5. Splendor	15. Pulsar
6. Wego	16. Samurai
7. Dazzler	17. RX
8. Unicorn	18. Shogun
9. Passion plus	19. Twister
10. Thunderbird	20. KTM

KRISHNAKUMAR K

I B.Sc. (Information Technology)

MATH PUZZLES

1. I am a number with a couple of friends, quarter a dozen, and you'll find me again. What am I?

Ans: **three**

2. Add me to myself and multiply by 4. Divide me by 8 and you will have me once more. What number am I?

Ans: **Any number.**

3. There is a three digit number. The second digit is four times as big as the third digit, while the first digit is three less than the second digit. What is the number?

Ans: **141**

4. In two years I know, I'll be twice as old as five years ago, said Tom. How old is Tom?

Ans: **12**

5. Four years ago, Alex was twice as old as Jake. Four years from now, Jake will be 3/4 of Alex's age. How old is Alex?

   Ans: **12**

PRAKASH S

I B.Sc. (InformationTechnology)

I think it would be a tragic statement of the universe if java was the last language that swept through".

-James Gosling.