Research Topics

# Recent Advances in Music Signal Processing

# IEEE SPS Volume 36 January 2019

## Automatic Music Transcription

* Investigates methods for transcribing polyphonic music produced by voice and pitched instruments.
* Music education, music creation, production searching and recommendation algorithms and musicology.
* Provides the main link between music signal processing and music language modelling.
* AMT is the musical equivalent of ASR, in that they both benefit from language modelling components in combination with processing acoustic components.
* **Key Challenges –**
* Inferring musical attributes from the mixture signal is an extremely underdetermined problem.
* Consonant intervals contain a lot of harmonic overlap in frequency making the separation even more difficult.
* There is no statistical independence between sources due to the synchronization of onsets and offsets between the performing musicians.
* Annotation of ground-truth transcriptions for polyphonic music is time consuming, subjective and requires high-expertise. This has limited supervised learning as musical scores can be viewed as weak labels.
* There are often a multitude of versions of any given musical work.
* **Future work –**
* Music Language Models - Extend RNN applications beyond note sequence recognition to chord mapping.
* Score informed transcriptions – specifically lead sheet informed transcription. The score/music piece is used as a prior to the performance to better inform the transcription.
* Context-specific transcription – Prior knowledge of instruments and recording environment to enhance transcription accuracy
* Expressive pitch and timing – Western notation conceptualize music as sequences of unchanging pitches being maintained for regular durations. There is no suitable notation which exists for performed singing/ or instruments without fixed pitch.
* Percussion and unpitched sounds – non-pitched sound classification in percussive instruments are particularly challenging due to the range of concurrent spectral profiles in the mixture.
* Evaluation metrics – creation of perceptually relevant evaluation metrics for AMT and the creation of evaluation metrics for notation level transcription remains an open problem

## Musical Source Separation

* Task of recovering one or more of the source signals from a mixture signal.
* Can be difficult when given access to the final recording mix after a host of modifications.
* **Future work –**
* Reduction of audible artefacts produced by most algorithms through the use of: phase retrieval techniques to estimate the phase of the target source, feature representations that better match human perception, MSS systems that model the signal directly in the time domain as waveforms.
* Improving the flexibility of MSS systems to deal with the richness of musical data, for example the separation of sources in a violin ensemble
* Listening tests are the only quality evaluation procedure to date.
* How can DNN based techniques be exploited for music separation? How can we avoid overfitting? There is not much data so you are likely to overfit!

## Deep Learning for Audio-Based Music Classification and Tagging

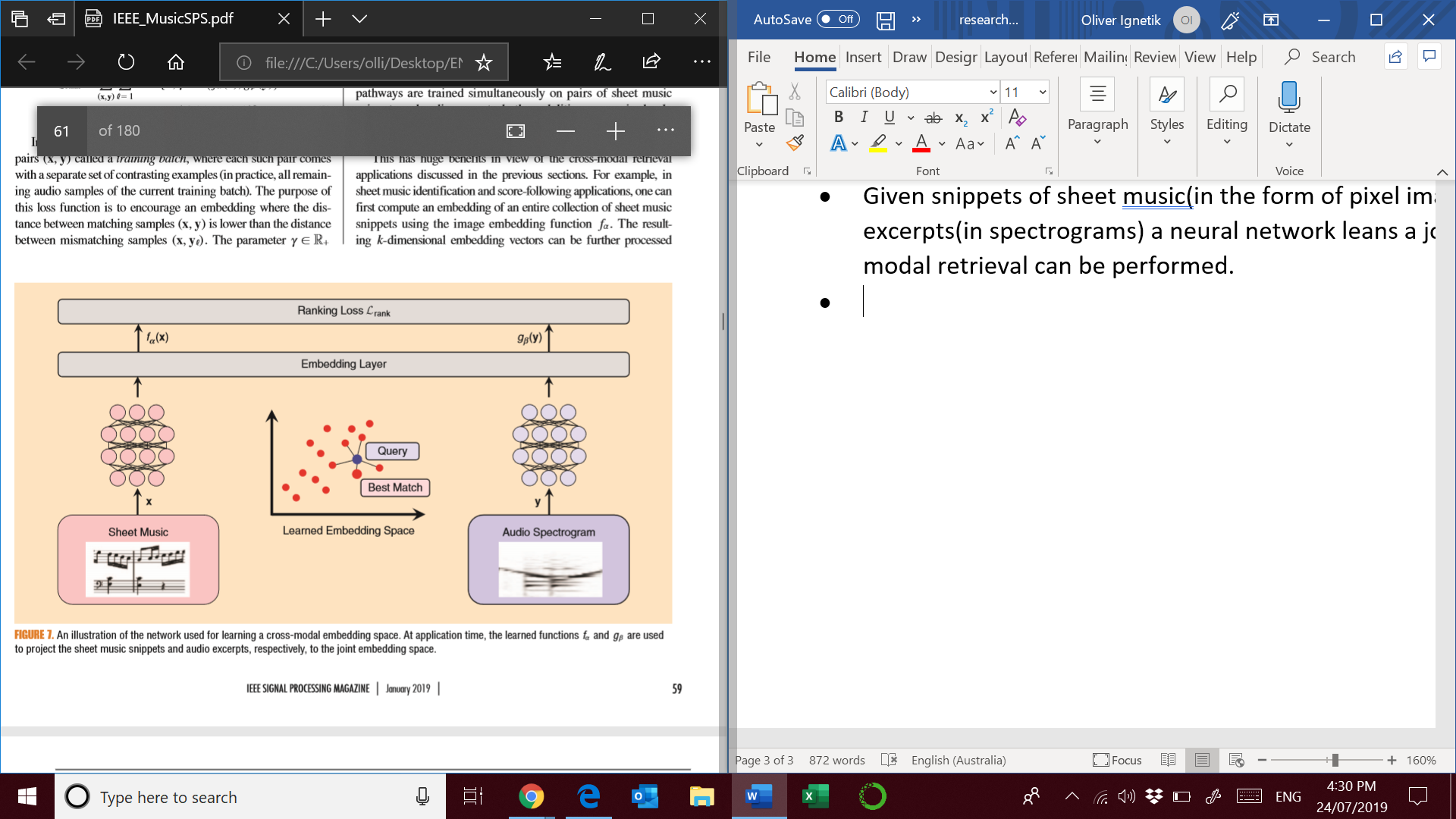
* How do we find songs that fit users’ tastes?
* Collaborative filtering based on play history and song rating, is hampered by popularity bias and the cold-start problem.
* Music Information retrieval (MIR) is centred on finding ways to automate the process of classifying music genre and mood and tagging music.
* Translating musical and acoustic features into numerical representations is the essence of music tagging.
* Possible applications include predicting hit songs and recommendations
* **Future work –**
* Copyright infringement limits widespread research on music tagging. Overfitting is inherent due to lack of data
* Music domain knowledge is needed in the design of networks. It might be beneficial to inform the network of midlevel features such as the use of different chords and syncopation.
* This could be achieved by including the musical scores in a multi modal approach. Little work has been done to use a score jointly with in a neural network.
* Little work has been done to jointly tackle the problem of source separation and tagging under one framework.
* Increase the diversity and coverage of labels by the possible use of social platform tags.
* How does one personalize music mood classification?

## Making Music More Accessible for Cochlear Implant Listeners

* Key musical features like pitch and timbre are poorly transmitted by CIs
* The most salient element of music are dynamics, pitch, rhythm and timbre
* Currently approximately 360 million people worldwide suffer from hearing loss
* Transmission of information between the auditory nerve and auditory periphery is interrupted
* Cochlear implants use electrodes to simulate this nerve however struggle to provide listeners with satisfactory perception of music
* This is due to a number of factors including: a low number of electrodes, spread of electrical fields in the cochlea causing broad excitation patterns and undesired channel interactions and restrictions of CI in transmitting fine structure of acoustic signals
* Solo instrumentation is used preferred as is pop and genres with high regularity
* **Future work –**
* Investigation to evaluation of music perception in live concerts by CI and normal-hearing listeners to help design new signal processing algorithms and sound coding strategies for listening
* Several sound coding strategies aim to improve pitch perception and melody recognition by enhancing transmission of temporal aspects of the input sound in the apical part of the cochlea
* Increase stimulation points beyond those provided by the electrodes
* Signal processing algorithms that reduce the complexity of music using DRNNS or NMF to remix music with clear vocals and enhanced melodies
* Further work should aim at unification of pre-processing and sound coding to improve their efficiency whilst improving latency and reducing complexity
* Testing in more realistic environments using virtual reality labs can be used to assess the perception of music in users.

## Cross-Modal Music Retrieval and Applications

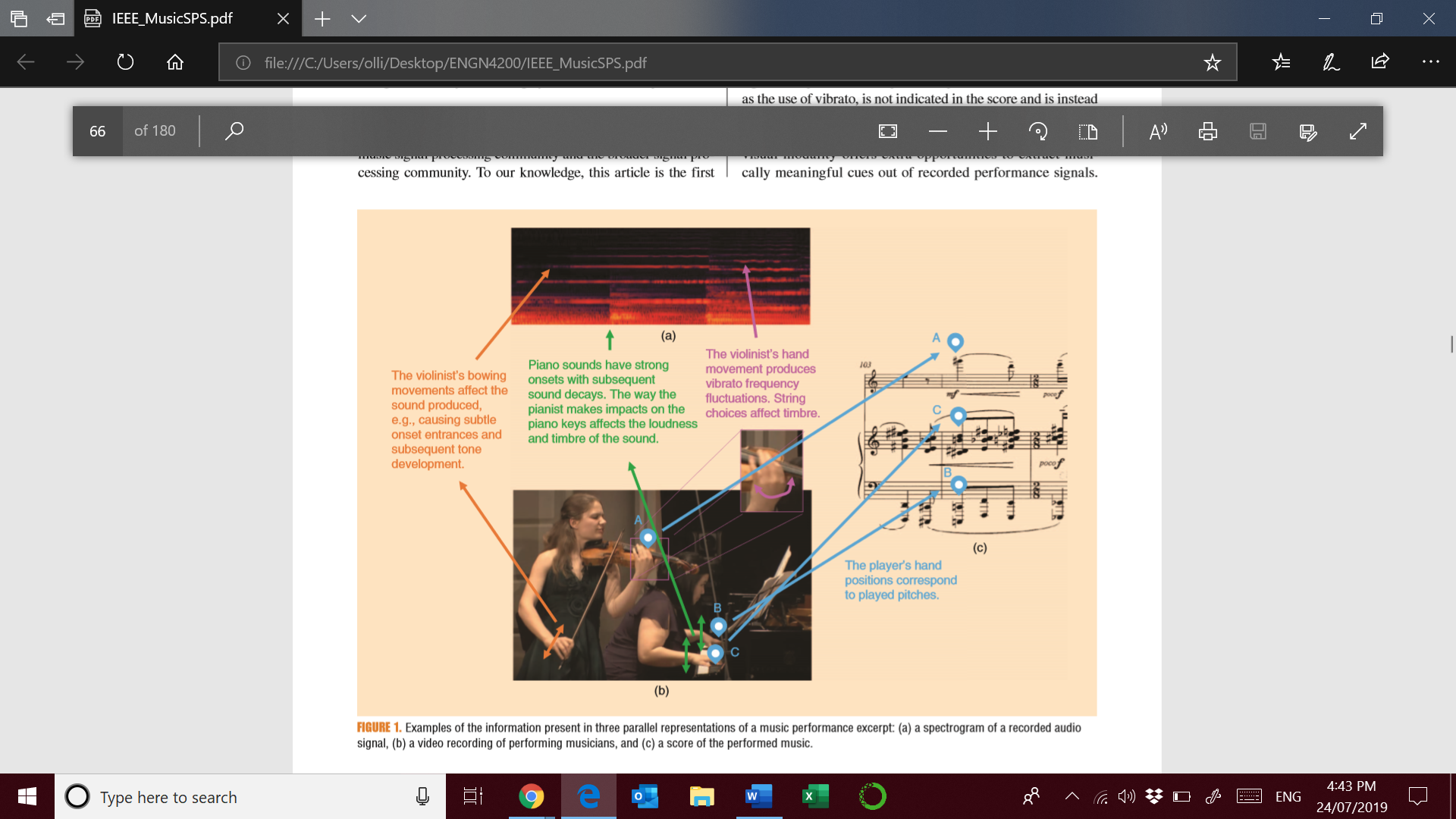
* Given a query in one modality find corresponding information and entities in other modalities
* This goes beyond exact audio identification like Shazam.
* There are approaches where sheet music and audio representations are converted into midlevel feature representation that capture musical properties related to pitches and harmony. They are then compared using alignment algorithms.
* Given snippets of sheet music(in the form of pixel images) and corresponding audio excerpts(in spectrograms) a neural network leans a joint embedding space on which cross-modal retrieval can be performed.



* **Future work –**
* Score following where the computer listens to a live performance and tries to read along with sheet music
* Piano Music Companion identifies live performances of live classical music in real time <https://www.youtube.com/watch?v=SUBtND_MJZs>
* Possible extension to a complete companion for whatever instrumentation or any medium.
* YouTube videos can be enriched with automatically generated musical annotations.

## Audio-visual Analysis of Music Performances

* Music performances are multimodal as visual aspects play an important role in expressing ideas and emotions.



* This is applicable to music tagging AMT and music source separation. Essentially more features are available for comparison.
* **Future work –**
* Music scenes are cluttered, meaningful motions are subtle
* Integration of audio and visual processing in the modelling stage of musical scene analysis.
* Lack of large scale annotated ground truths due to copyright laws.
* AMT can benefit by detection of play/non play activity.
* Source separation will greatly benefit, motion of players is highly correlated to sound sources.
* Deep learning has been utilized for vision based detection of acoustic timed events.

## Music Interfaces Based on Automatic Music Signal Analysis

* Music listening interfaces have remained unchanged except for a continuous playback slider.
* Music interfaces based on machine listening allow users to change the playback to logical positions. RefraiD is a commonly used SP method to identified choruses and different sections with high similarity.
* Align existing notation to audio (score to audio notation)
* This can be achieved by converting a piece to feature sequences and using DTW or hidden Markov models to align them.
* “Intelligent editors” can automate and simplify routine edits in music production.
* Possible interface that can accept or reject the computers suggested changes.
* **Future Work –**
* Improvements in AMT and source separation will help to improve the granularity of control in these interfaces and improve music education. For example these approaches will allow for filtering methods other then equalization, bass and treble.
* Interface that combines lead sheet for jazz with multimodal interface.
* Songle provides a crowdsourcing interface that encourages users to correct errors by selecting from an alternative list of annotations.

## An Introduction to Signal Processing for Singing-Voice Analysis

* Autotune essentially
* Manipulation of intermediary representations between symbolic information and audio signal representations.
* High level content attributes like emotions, vocal fit and lyricism can be combined with recommendation approaches to better facilitate playlist generation etc.
* Singer ID to resolve metadata issues
* Melody estimation – determining the pitch of the singing voice.
* Audio lyrics alignment, lyrics transcription
* **Future Work –**
* Vocal activity detection – obtaining data for training and evaluation is hampered by copyright laws
* Singer ID – It is possible to collect large training data sets. Combination of modern source separation can help with ID of singers.
* Language identification – FMA contains non-English tags for several hundred recordings
* Subjective evaluation of signing voice models remains a challenge due to the reliance on human participation.

## Speech to singing voice conversion

* Strong commercial drive since it can beautify the signing renditions of amateur singers.
* It can be used to automatically generate reference singing for vocal learners and help with
* Deep learning approach to STS conversion to help with temporal alignment and parameter conversion
* Speech and singing vocals are produced by the same human voice and share characteristics. However, they do manifest through different acoustic characteristics.
* STS conversion must address several research problems including the temporal alignment between two very different signals.
* Mapping of dissimilar acoustic characteristics and preserving fine attributes of speakers during conversation.
* **Future work –**
* Improve the quality of temporal alignment and parameter conversion
* Real time implementation of the algorithms that are not well studied.
* Reducing the artefacts of introduced by individual modules in the SPS pipeline.

## Model-Based Digital Pianos

* Most commercial digital pianos are based on sample playback, however the sound can be reproduced by physically modelling a piano.
* Finding numerically optimized signal processing models that allow sound synthesis in real time based on the users inputs
* This has many applications including the development of new pianos.
* It is now possible to sample each note with various key velocities/loudness levels
* However there are a number of phenomena that cannot be concisely reproduced by recorded samples including: free vibration of the strings when the sustain pedal is pressed, the coupling between the strings of the sounding notes, or restrike of an already sounding string, continuous alteration of pianos by tuning or changing the hardness of the hammer
* **Future work –**
* The sounds produced are still disappointing despite physical modelling attempts this is due both to absence of certain features in the models and inaccuracies.
* Digital waveguide and modal based piano models are the most prominent models that have been investigated
* Future research could entail understanding the string coupling at the bridge, the effect of the crown on soundboard vibrations or finding a way to better model the shock of the key on the structure.

# Selected topics on music SPS

## Automatic mood detection and tracking of music audio signals

Published 2006 January

<https://ieeexplore.ieee.org/document/1561259>

* Automated mood detection from acoustic music data
* Automation based on Western psychological theories
* Three features: intensity, timbre and rhythm
* Breaking up music into homogenous emotional expressions

## Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition

Published 2012

<https://ieeexplore.ieee.org/document/5740583>

* Compares context dependant (CD) hidden Markov model and a deep neural network in speech recognition
* Investigates pre-training algorithm known as deep belief network.

## Understanding Effects of Subjectivity in Measuring Chord Estimation Accuracy

Published 2013

<https://ieeexplore.ieee.org/document/6587770>

* Automatic Chord Estimation in Music Information Retrieval Evaluation eXchange task with a major/minor chord alphabet
* Comparison of reference annotation to the ground truth annotation.
* Ground truth annotation is generated by a musically skilled individual.

## Acoustic Modelling Using Deep Belief Networks

Published 2011

<https://ieeexplore.ieee.org/document/5704567>

* Typical ASR systems use hidden Markov models to model the structure of speech signals.
* Window of 25ms of speech, is analysed by identifying a set of distinctive feature landmarks
* Feedforward neural networks compared to GMMs
* First application to acoustic modelling of neural networks in which multiple layers of features are generatively pre-trained.

# IEEE SPS Volume 36 September 2018

# New Foundation

* Device size and power consumption are two key challenges to the IoT. The IoT will soon require the ability to store an unprecedented amount of data, resulting in astronomical energy costs.
* CMOS technology will soon be unable to undergo further miniaturization.
* Memristors are superior and viable replacements to replacing transistors. They can retain memory of its last resistance state after being switched off.
* They have applications in Neuromorphic computing as they can change their resistance states continuously.

# Practical Backscatter Communication Systems for Battery-Free Internet of Things

* Ultralow-power wireless communication paradigm
* The ability to offer submilliwat power consumption makes it competitive core technology for IoT applications
* Similar to a heliograph utilizing a reflection while manipulating schema.
* A special device called a backscatter tag reflects the incoming excitation signal emitted by nearby (carrier) transmitters. At the same time it selectively changes the amplitude, frequency and/or phase of the signal for modulation
* **Future work –**
* Advanced modulation schemes – OFDM backscatter throughput improvements
* Downlink and full duplex – Investigate the tradeoff between persistence in the excitation signal for uplink and intermittent patterns of the signal during downlink
* MIMO – Beamforming can effectively help the backscatter tag to get a strong excitation signal

IoT Security Techniques Based on Machine Learning

* IoT has to protect users against spoofing attacks, denial of service, jamming and eavesdropping
* Most existing security solutions execute heavy load on devices. The more primitive the device the more vulnerable they are to attack.
* There are a host of ML techniques such as K-NN, SVMs and NNs used in building defence mechanisms
* **Future work –**
* Transfer learning to accelerate the beginning of the learning process through data mining to reduce random exploration
* Reduction of computation and communication overhead. Investigate dFW have to be investigated especially for scenarios with no access to cloud-based servers

# Intelligent Signal Processing and Coordination for the Adaptive Smart Grid: An Overview of Data-Driven Grid Management

* Heirarchical signal processing and actuation framework that can enable all encompassing coordination of thousands of actuating power entities to maintain efficiency accounting for infrastructure limits.
* Amid the rapidly changing landscape of electricity supply and demand upgrades to the underlying power system are falling behind.
* Cyber enabled sensors and actuators can be used to enable adaptive monitoring by capitalizing on advances in data analytics and intelligent signal processing techniques.
* Extensive opportunities lie in the effective use of vast volumes of grid-monitoring data generated every second by measuring devices. Designing signals that capture general trends in the power grid can help with grid operations.
* Information about the global state of the system via millions of nodes in the system is a great opportunity for data analytics

# Android Malware detection using a deep learning method

<https://ieeexplore-ieee-org.virtual.anu.edu.au/stamp/stamp.jsp?tp=&arnumber=8443370>

* With the widespread use of smartphones malware attacks of increased exponentially
* Various features are used to reflect the properties of android applications.
* In 2017 there were 750 000 new malware attacks
* Android devices allow users to install third party software
* Static analysis is vulnerable to code encryption or packing

# Dr Sadeghi research

<http://users.cecs.anu.edu.au/~parastoo/research.html>

# Information-theoretic data privacy

Dr. Sadeghi

* Poor anonymization or de-identification methods for sensitive data
* Inference attacks work when logical inferences can change statistical averages amongst users in a database
* The privacy protection method should be invariant to the available computational power
* Understanding of information theory can help us use optimization techniques to strike the best quality between anonymity and utility of data
* Information theory has the power to allow a curator to dial up or dial down the access arrangements according to privacy risks and fidelity requirements.
* **Future work –**
* How can we search for an optimal soft transition over the probability simplex?
* Investigate better MDSF algorithms to improve the performance of IAC-MDSF algorithm

<https://arxiv.org/pdf/1901.06629.pdf>

# Index Coding

* Problem entailing the broadcast of a set of n messages from a sever to receivers with side information.
* Distributed index coding involves more then one server which can contain a subset of messages and has applications in distributed storage/broadcast networks and vehicular networks like in the defence force

<http://users.cecs.anu.edu.au/~parastoo/papers/2016/ITW_2016.pdf>

# Network Coding

* Contributes to understating fundamental performance limits of wired and wireless data networks
* Network coding enhances throughput performance but oftentimes at the cost of decoding latency
* For example in generation-based random linear network coding all packets can only be decoded after the full transmission block is received.

<http://users.cecs.anu.edu.au/~parastoo/papers/2015/NETCOD_2015_3.pdf>

<http://users.cecs.anu.edu.au/~parastoo/papers/2018/ITW_2018_3.pdf>