AUSTRALIAN RESEARCH COUNCIL
Discovery Projects
Proposal for Funding Commencing in 2018

DP

PROJECT ID: DP180103563

First Investigator: Prof Hua Wang

Admin Org: Victoria University

Total number of sheets contained in this Proposal: 71

Information on this form and its attachments is collected in order to make recommendations to the Minister on the allocation of financial assistance under the Australian Research Council Act 2001 and for post award reporting. The information collected may be passed to third parties for assessment purposes. It may also be passed to the National Health and Medical Research Council and any other Australian Government Department or Agency for the purpose of checking eligibility. In other instances, information contained in this Proposal can be disclosed without your consent where authorised or required by law.

Certification

Certification by the Deputy/Pro Vice-Chancellor (Research) or their delegate or equivalent in the Administering Organisation

I certify that—

- I have read, understood and complied with the ARC *Funding Rules for schemes under the Discovery Programme (2016 edition)* (the Funding Rules) and to the best of my knowledge all details provided in this Proposal form and in any supporting documentation are true and complete in accordance with these Funding Rules.
- Proper enquiries have been made and I am satisfied that the Participants and the organisations listed in this Proposal meet the requirements specified in the Funding Rules. I will notify the ARC if there are changes to any named Participant or organisation after the submission of this Proposal.
- To the best of my knowledge, all Conflicts of Interest relating to parties involved in or associated with this Proposal have been disclosed to this Administering Organisation, and, if the Proposal is successful, I agree to manage all Conflicts of Interest relating to this Proposal in accordance with the *Australian Code for the Responsible Conduct of Research* (2007).
- The listed Participants are responsible for the authorship and intellectual content of this Proposal, and have appropriately cited sources and acknowledged significant contributions to this Proposal.
- I have obtained the agreement, attested to by written evidence, of all the relevant Participants and organisations necessary to allow the Project to proceed. This written evidence has been retained and will be provided to the ARC if requested.
- This Proposal complies with the eligible research requirements set out in the *ARC Medical Research Policy* located on the ARC website.
- This Proposal does not duplicate Commonwealth-funded research including that undertaken in a Commonwealth-funded Research Centre.
- If this Proposal is successful, I am prepared to have the Project carried out as set out in this Proposal and agree to abide by the terms and conditions of the Funding Rules and the ARC *Funding Agreement regarding funding for schemes under the Discovery Programme (2016 edition).*
- The Project can be accommodated within the general facilities in this organisation and, if applicable, within the facilities of other relevant organisations specified in this Proposal, and sufficient working and office space is available for any proposed additional staff.
- All funds for this Project will only be spent for the purpose for which they are provided.
- The Project will not be permitted to commence until appropriate ethical clearance(s) has/have been obtained and all statutory requirements have been met.
- I consent, on behalf of all the parties, to this Proposal being referred to third parties, who will remain anonymous, for assessment purposes.
- I consent, on behalf of all the parties, to the ARC copying, modifying and otherwise dealing with information contained in this Proposal.
- To the best of my knowledge, the Privacy Notice appearing at the top of this form has been drawn to the attention of all the Participants whose personal details have been provided at the Personnel section.

Part A - Administrative Summary (DP180103563)

A1. Proposal Working Title

(Provide a short working title of no more than 75 characters (approximately ten words). Avoid the use of acronyms, quotation marks and upper case characters.)

Increasing data quality with group associations in outsourcing environment

A2. Person Participant Summary

(Add all people participating in this Proposal as a Chief Investigator or Partner Investigator. Note that a person's RMS email address must be used to invite them to participate in this Proposal. Click on the information icon or refer to the Instructions to Applicants for further information.)

Number	Name	Participant Type	Current Organisation(s)	Relevant Organisation
1	Prof Hua Wang	Chief Investigator	Victoria University	Victoria University
2	Prof Yanchun Zhang	Chief Investigator	Victoria University	Victoria University
3	Dr Jinli Cao	Chief Investigator	La Trobe University	La Trobe University

A3. Organisation Participant Summary

(Add all organisations participating in this Proposal. Click on the information icon or refer to the Instructions to Applicants for further information.)

Number	Name	Participant Type
1	Victoria University	Administering Organisation
2	La Trobe University	Other Eligible Organisation

A4. Proposal Summary

(Provide a written Proposal summary of no more than 750 characters (approximately 100 words) focusing on the aims, significance, expected outcomes, benefits and impacts of this Project. Avoid the use of acronyms, quotation marks and upper case characters. Please click on the information icon or refer to the Instructions to Applicants for further information.)

Outsourcing of data storage is increasingly common, but poses major problems for data utility and confidentiality. This project aims to discover how tuples in fragments can be grouped for producing loose associations so to increase the utility of queries executed over fragments. It creates a framework to perform grouping and produce loose associations that satisfy information protection goals while achieving utility for queries. The outcomes of the project are to formally specify and develop a minimal fragmentation model to validate loose association rules while minimising data leakage risks. The outcomes will benefit Australians through enabling sharing and linking increasingly large datasets securely and cheaply.

A5. Impact Statement

(In no more than 500 characters (approximately 75 words), outline the intended impact of the Project. Click on the information icon or refer to the Instructions to Applicants for further information.)

This project addresses vital problems associated with increasing data utility in outsourcing environments. The developed algorithms and techniques will be implemented in freely available open-source prototype software, allowing Australian scientists in demography, criminology and other application areas to share their outsourced datasets securely and inexpensively.

Part B - Classifications and Other Statistical Information (DP180103563)

B1. Does this Proposal fall within one of the Science and Research Priorities?

Yes

Science and Research Priority	Practical Research Challenge
Resources	Knowledge of environmental issues associated with resource extraction.

B2. Field of Research (FoR)

(Select up to three classification codes that relate to your Proposal. Note that the percentages must total 100%. Click on the information icon for further information.)

Code	Percentage
080109 - Pattern Recognition and Data Mining	60
080605 - Decision Support and Group Support Systems	40

B3. Socio-Economic Objective (SEO-08)

(Select up to three classification codes that relate to your proposal. Note that the percentages must total 100%. Click on the information icon for further information.)

Code	Percentage
890301 - Electronic Information Storage and Retrieval Services	100

B4. Interdisciplinary Research

(This is a 'Yes' or 'No' question. If you select 'Yes' two additional questions will be enabled:
1. Specify the ways in which the research is interdisciplinary by selecting one or more of the options below.
2. In no more than 375 characters (approximately 50 words), indicate the nature of the interdisciplinary research involved.)

Does this Proposal involve interdisciplinary research?

No

Please specify the ways in which the research is interdisciplinary by selecting one or more of the options below.

In no more than 375 characters (approximately 50 words), please indicate the nature of the interdisciplinary research involved.

B5. Does the proposed research involve international collaboration?

(This is a 'Yes' or 'No' question. If you select 'Yes' two additional questions will be enabled:
1. Specify the nature of the proposed international collaboration by selecting one or more of the options below.
2. Specify the countries which are involved in the international collaboration.)

No

B6. What is the nature of the proposed international collaboration activities?

(Select all options from the drop down list which apply to this proposal by clicking on the 'Add' button each time you

select an option.)

B7. If the proposed research involves international collaboration, specify the country/ies involved

(Commence typing in the search box and select from the drop-down list the name of the country/ies of collaborators who will be involved in the proposed Project. Note that Australia is not to be listed and is not available to be selected from the drop-down list.)

B8. How many PhD, Masters and Honours places will be filled as a result of this project?

(The ARC is interested in reporting the number of Research Students that would be involved in this Proposal if it is funded. Enter the number of student places (full-time equivalent) that will be filled as a result of this project:)
Number of Research Student Places (FTE) - PhD

2

Number of Research Student Places (FTE) - Masters

Number of Research Student Places (FTE) - Honours

Part C - Project Description (DP180103563)

C1. Project Description

(Please upload a Project Description as detailed in the Instructions to Applicants in no more than eight A4 pages and in the required format.)

Uploaded PDF file follows on next page.

PROJECT TITLE

Increasing data quality with group associations in data outsourcing environments

AIMS AND BACKGROUND

This project aims to develop a fragmentation-based framework to protect confidentiality and satisfy utility requirements when outsourcing data to external storage. The framework relies on tuples in fragments that can be grouped for producing loose associations so to increase the utility of queries executed over fragments. Privacy risk models that rely on the framework will be developed with comparisons to established technical analysis of loose associations satisfying a given level of protection for sensitive associations, while achieving utility for queries over different fragments (Tasks 1, 2). A multiple loose association for data publishing is defined as taking into account an arbitrary number of fragments and including the properties that need to be guaranteed to ensure that a loose association satisfies a given privacy degree and modification algorithms (Tasks 2, 3). An innovative logical language for the utility of loose associations in terms of providing better response to queries is proposed as the basic technique within this framework (Task 1). Models will be implemented to demonstrate the feasibility of the framework and generalize loose associations that operate over an arbitrary number of fragments (Tasks 4 and 5). Finally, we will evaluate the performance and the efficiency of approaches over both synthetic and real data sets (Task 6).

Background

Outsourcing data storage to the cloud and/or commercial operators has the benefits of convenience and reduced cost, but data owners are increasingly concerned about the confidentiality of their sensitive information [Yu et al. 2014, Wang, Zhang and Cao. 2009]. The challenge of protecting the confidentiality of private information has received much attention from the research and development communities, and many approaches have been proposed [Jhawar et al. 2012, Huang et al. 2017]. Traditional methods for confidentiality involve an encryption layer; query evaluation over encrypted data requires adoption of either indexing techniques [Kohler et al, 2015] or specific encryption approaches such as homomorphic encryption, which provide only limited capabilities for querying. The need to maintain data in a form that provides better support for queries is one of the key drivers of the search for solutions relying on fragmentation for protecting sensitive associations in data outsourcing [Berman and Raskhodnikova, 2014, De Capitani di Vimercati et al. 2014]. Fragmentation protects sensitive associations among data by splitting them in different fragments (vertical data views) that are not linkable. The advantage of fragmentation over data encryption is the ability to query actual data (in contrast to the indexes used when data are completely encrypted) and therefore provide more convenience in terms of data accessibility and query performance. Fragmentation also represents a useful paradigm for enforcing protection requirements and producing different views over data that can be publicly released without the risk of disclosing sensitive information. This project will address the disclosure problem of outsourcing data within a fragmentation-based framework.

Enhancing data utility is the other side of confidentiality, and is also a major concern for data storage users. Loose associations are applied to data utility and have been recently proposed as a complement to fragmentation [De Capitani di Vimercati, et al. 2015]. In fact, providing complete protection to sensitive associations can, in many cases, be considered overly conservative and smaller protection guarantees (meaning uncertainty about the associations among data) are often acceptable. Loose associations complement data fragmentation by providing some information on the association among tuples in different fragments. Intuitively, being fragments unlinkable without loose associations, a tuple in a fragment could have, as its corresponding tuple, any tuple in another fragment. Loose associations provide information on the relationships between tuples of different fragments at the granularity of groups of tuples, thus maintaining some degree of protection over the associations. The original definition of loose associations assumes that a fragmentation includes two fragments only and a single loose association is defined between this pair of fragments. A fragmentation may, however, include an arbitrary number of fragments. The presence of multiple loose associations may unfortunately open the door to privacy breaches. In fact, while the associations released in loose form are protected, the publication of multiple loose associations could indirectly expose other sensitive associations since recipients could be able to reconstruct them.

In summary, privacy preserving and data utility in data publishing environment faces three challenges, as detailed below.

Challenge 1. Loose associations on an arbitrary number of fragments: The data owner can specify multiple associations among different pairs of fragments with the assurance that their combination cannot introduce leakages. Also, it allows specifying loose associations involving more than two fragments. The challenge is how to define a loose association model encompassing all the fragments and all the confidentiality constraints [Samarati, 2014, Sun, et al. 2016]. The model can provide a loose association involving a different subset of fragments and is guaranteed to maintain the aimed degree of protection to the sensitive information.

Challenge 2. Data sharing in data publishing environment: The strong need for sharing and disseminating information that characterizes our global internet worked society raises several privacy concerns (e.g., [De Capitani di Vimercati et al. 2015, Gentry and Halevi, 2011, Tang et al. 2016]). In fact, the vast amount of data collected and maintained in the digital infrastructure often includes sensitive information that must be adequately protected. There is then a clear trade-off between data utility, such as the need to easily access, use, and distribute information, and the equally strong need to provide proper protection guarantees to sensitive information. Traditional solutions aimed at protecting data undergoing public or semi-public release are based on k-anonymity and differential privacy [Dwork, 2006], which protect respondents identities and their sensitive information by releasing a sanitized version of the data. However, these solutions are not applicable in scenarios characterized by the need to release non-modified information.

Challenge 3. Enhancing data utility with query: An important aspect to be taken into consideration in the development and application of protection techniques for ensuring confidentiality of sensitive information is the need to maintain utility in the data while avoiding their over protection, where utility encompasses both the availability of certain information as well as the ability to perform queries over the data [Jiang et al. 2016]. The challenge lies in the analysis of data utility with advanced query algorithms and the optimal methods to reduce the complexity.

Existing fragmentation and privacy preserving approaches have limitations related to disclosure to third parties (Challenges 1 and 2) and efficient data utility solutions (Challenge 3).

Aims

This project investigates novel approaches to loose associations and data utility in data publishing environments. The basic idea behind the approach is to specify multiple associations among different pairs of fragments with the assurance that their combination cannot introduce leakages. The project presents a formal model for universal loose associations, encompassing all the fragments and all the confidentiality constraints. It will allow the data owner to outsource, based on data utility requirements, and intelligently manage the loose associations involving a different subset of fragments, thus providing the aimed protection level to the sensitive associations.

We will develop a fragment involved approach that takes into account queries to be executed so to build loose associations that enhance utility for them. We will then produce a heuristic algorithm for the computation of a loose association, and conduct an extensive experimental analysis over synthetic and real data sets aimed at evaluating the efficiency, efficacy and scalability of the algorithm, as well as the utility of the computed loose association.

Aim 1. To develop a framework to recognise and standardise group associations to enhance data utility for outsourcing data sets. (Challenge 1)

Aim 2. To provide novel optimised algorithms for the computation of a loose association. (Challenge 2)

Aim 3. To efficiently discover access query algorithms by using the original data sets and newly generated data sets. (Challenge 3)

Aim 4. To implement a prototype set of loose association rules over synthetic and real data sets aimed at evaluating the efficiency, efficacy and scalability of the developed algorithm, as well as the utility of the computed loose association. (Challenges 1, 2 and 3)

The importance of data publishing and data utility has been recognised for a long time [Lin and Chen, 2011, Kabir et al. 2015]. However, the concept of loose association has not been supported in existing privacy preserving models, especially in data publishing environments. This project will bridge the gap between private information protection technology and association models in outsourcing environments.

INVESTIGATORS

Professor Hua Wang will contribute 20% of his time to the project to manage and take overall responsibility for the project and lead the research on the loose association model (Tasks1 and 3), the optimal fragmentation algorithm (Task 3), and supervise the Research Assistant and PhD students in system design, evaluation and implementation (Tasks 4, 5 and 6). His creativity and insight are crucial to the research program.

Professor Yanchun Zhang will make 20% of his time to work on the loose association model (Task 1), intelligent key distribution and standardization (Task 2) and supervise the Research Assistant and PhD students in system design and evaluation (Tasks 3, 4, 6).

Dr Jinli Cao will contribute 20% of her time for the project to develop improved data mining algorithms and techniques for intelligent key distribution and standardization (Task 1), improved access policy techniques in outsourced database systems (Task 3), and jointly supervise the PhD student and the *RAs* (in system design and implementation) (Tasks 4, 5 and 6). Her expertise will help overcome the formidable technical difficulties in these advanced applications.

PhD student 1: will work on data outsourcing model with privacy risk and confidential constraints and minimal fragmentation methods, design SQL query model, the relationship between fragmentation and utility approaches (Tasks 1 and 3). Additionally, the student is required to work on the demonstration of the access queries and validations in real data sets (Task 5).

PhD student 2: will involve the analysis and development of the fragment-based intelligent algorithms (Task 2), design a prototype of the fragment-based framework for a public data set to demonstrate the feasibility of the proposed k-grouping methods (Task 4); develop graphical user interfaces to handle confidential constraints and utility requirements (Task 6).

Research assistant: will have expertise in data mining and privacy preserving models and be primarily responsible for using different fragmentation methods to create, investigate and compare various loose association approaches in real data sets; extending application of the loose associations and forcing onto the data set in outsourced environments. The *RA* will also implement and evaluate optimal methods and data utility requirements for experiments and evaluation (Tasks 2, 3, 4, 5 and 6).

PROJECT QUALITY AND INNOVATION

Significance

This project involves theoretical development and a practical demonstration of how to enhance data quality and protect sensitive information and associations in outsourced data systems.

Significance 1. The proposed loose association methods will achieve both data utility and efficiency for users and sensitive data protection for owners in outsourced environments. Consider the following real-life scenario: A hospital has a patient dataset containing details such as name, date of birth, education, income and so

on (Figure 1 (a)). Health institutes need patient data to establish causes of death, study patterns of disease and make predictions for other patients. This scenario poses several challenges: (1) Patients do not wish to publish all their information, especially sensitive data, so the dataset has to be modified before release. (2) Health institutes obviously require useful data and some items are essential, hence the modification must be minimal. Health institutes are not allowed to access information other than what is being queried in a dataset. This is an important aspect in the real world, because the institute may have to pay for the dataset and the hospital must provide only the data that has been requested. (3) How to design content-level fine-grained access methods for clients? This challenge is very hard to solve, as it requires variable capabilities for different clients. This project focuses on these important challenges.

Name	YoB	Edu	ZIP	Disease	Race	Income
Alice	1974	B.Sc	90015	Flu	Asian	5K
Bob	1965	MBA	90038	Diabetis	White	2K
Carol	1976	Ph.D	90001	Calculi	Black	4K
Greg	1975	M.Sc	90025	Flu	Black	3K

a. Patients data

c1={YoB, Edu}
c2={ZIP, Job}
c3={Name, Disease}
c4={YoB, ZIP, Disease}
c5={YoB, ZIP, MarStatus}

b. Confidential constraints

Figure 1: An example of patient relation, confidential constraints over (a)

$v_1 = Name \vee ZIP$
$v_2 = \{Edu \wedge ZIP\} \vee SSN$

v. Utility requirements

Name	YoB
Alice	1974
Bob	1965
Carol	1976
David	1972

F_l

Edu	ZIP
B.Sc	90015
MBA	90038
Ph.D	90001
M.Sc	90025

F_r

Figure 2: Utility requirements and two fragments F_l and F_r

Name	YoB	G
Alice	1974	ny1
Bob	1965	ny1
Carol	1976	ny2
David	1972	ny2

$F_l + grouping$

Edu	ZIP	G
B.Sc	90015	ez1
MBA	90038	ez2
Ph.D	90001	ez1
M.Sc	90025	ez2

$F_r + grouping$

G_l	G_r
ny1	ez1
ny1	ez2
ny2	ez1
ny2	ez2

Loose association

Figure 3: An example of loose association

Significance 2. A generalised system for data utility will promote wide use of efficient data storage for various purposes. Much research has recently been dedicated to techniques of providing different forms of protection and computing a "sanitized" version of data for publication [Huang, Wang and Cao, 2016]. Recently, most of this work has focused on k-anonymity and its variations (e.g., l-diversity) for protecting respondents identities and their sensitive information when releasing microdata [Biskup, 2016, Berman and Raskhodnikova, 2014]. Although some approaches have addressed the problem of safeguarding the utility of sanitized data, the problem of considering utility requirements has not been investigated.

Significance 3. The developed methods for data utility are optimised and efficient. Most research on data protection assumes the data to be entirely encrypted, and focuses on the design of techniques for the efficient execution of queries. One of the first proposals towards the solution of this problem is presented in [Yu et al. 2014], where data is encrypted on the client side before being stored in the distrusted external serr. In order to answer all database queries, the client needs to obtain the entire database from the server, decrypt it, and execute the query on this decrypted database. This approach is too expensive to be practical and difficult to enforce in real environments due to its inefficiency.

Innovations

Innovation 1. The proposal extends existing research by presenting a complete framework for association rules via intelligent algorithms.

This project introduces the novel problem of a loose association model for sensitive information and how to decrease the precision of the background knowledge the data leakage may possess. The notion of privacy requires that, for each fragmentation, there are at least k distinct items that are indistinguishable from the item regarding their supports.

Innovation 2. This project presents an extended data utility model with minimal fragment technology to protect sensitive information. The methods to be designed in the project will be able to recover the association

patterns and their correct support. We propose that it will create and keep a compact structure. The new model also provides an efficient strategy for incrementally maintaining the synopsis against updates in the form of appends.

Innovation 3. We will conduct a formal analysis for the designed loose association methods The loose association method based on the data utility model will prove the probability that an individual item, a transaction or a pattern can be acquired and comprehended by an unauthorised third party can always be controlled by the owner by choosing the protection threshold k.

Framework and methods

Given a generic fragmentation F composed of an arbitrary number of fragments, different group associations can be published on different fragment pairs.

Models

We consider a scenario in which a data owner wishes to release her data for publication or external storage. Data, represented for convenience as a single relation s over relational schema $S(a_1, ..., a_m)$, are subject to confidentiality constraints stating that certain information (individual attributes or associations among them) is to be considered sensitive and should therefore not be disclosed. A confidentiality constraint is formally defined:

Confidentiality constraint Given a relation schema $S(a_1, ..., a_m)$, a confidentiality constraint c over S is a subset of the attributes $\{a_1, ..., a_m\}$ in S.

Confidentiality constraints are enforced before release by avoiding disclosure of sensitive attributes (singleton constraints), and fragmenting the relation into vertical views so to break sensitive attribute associations (non-singleton constraints). Note that non-singleton constraints can also be enforced by non-release of a subset of the attributes in the constraint. At the schema level, fragmentation splits then S into a set $F = \{F_1, ..., F_n\}$ of fragments. Each F_i corresponds, at the instance level, to the vertical view F_i obtained projecting s over F_i. Given a fragmentation $F = \{F_1, ..., F_n\}$, sensitive information modelled by a set C of confidentiality constraints is protected by ensuring that: i) no individual fragment $F \in F$ contains all the attributes involved in a confidentiality constraint (i.e., $\forall F \in F, \forall c \in C : c \nsubseteq F$); and ii) fragments are disjoint (i.e., $\forall F_i, F_j \in F, i \neq j : F_i \cap F_j = \phi$). We assume data to be fragmented no more than necessary to satisfy the constraints, and therefore fragmentations to be minimal, that is, merging any two fragments would violate at least one constraint. Figure 1 illustrates an example of a relation to be released, of confidentiality constraints over it, and of fragmentations F_l, F_r satisfying the constraints.

Utility requirements Given a relation schema $S(a_1, ..., a_n)$, a Utility requirement v over S is a monotonic boolean formula over $\{a_1, ..., a_n\}$.

In fact, in many real-world scenarios, the specification of confidentiality constraints on one side and the specification of desired views of data to be published on the other side belong to different authorities. Intuitively, utility requirements impose the inclusion, or joint inclusion, of attributes in fragments of a fragmentation. The semantics of a utility requirement are therefore easily explained with reference to fragments. As an example, Figure 1(c) reports possible utility requirements over patients data in Figure 1(a). Here, v_1 states that either names of patients or their ZIP codes should be released; v_2 states that either education and ZIP codes of patients in association should be released or the SSN of patients should be released.

Fragmentation may completely break the associations among attributes appearing in different fragments. In fact, since attributes are assumed to be independent, any tuple appearing in a fragment could have as its corresponding part any other tuple appearing in another fragment. In some cases, such protection can be overkill and a lower uncertainty on the association could be preferred to mitigate information loss. A way to achieve this is to publish an association among tuples in fragments at the level of groups of tuples (in contrast to individual tuples), where the cardinality of the groups impacts the uncertainty over the association, which therefore remains loose. Hence, group associations are based on grouping of tuples in fragments.

This project investigates loose associations and data utility for data publishing in outsourced environments, and is divided into the following six interrelated tasks.

Task 1: Advanced data publishing model

We propose a novel modelling of the problem of protecting privacy when publishing data that explicitly takes into consideration both privacy needs and utility requirements. Our setting of the privacy problem is generic and does not assume, like typical k-anonymity solutions, a preliminary definition of identifying, quasi-identifying and sensitive data. Rather, it supports the specification of *confidentiality constraints*, generically capturing privacy needs as sensitive attributes, or sensitive associations among them, that need to be protected. *Utility requirements* provide an explicit means for data publishers and/or recipients to express the fact that certain data need to be published. Utility requirements may come, for instance, from third parties (e.g., research institutions) to which the data are released. Also, utility requirements reflect the fact that certain data are already available (e.g., from other external sources), avoiding publication of data whose combination with those already available might compromise privacy. This project is based on fragmenting data to break associations among them, guaranteeing that both confidentiality constraints and utility requirements are respected.

The project also incorporates the idea of complementing fragments with *loose associations*. Intuitively, loose associations partition tuples within fragments in different groups and release association information at the group level, as opposed to releasing the actual tuple-to-tuple association. The loose association problem is characterized by a privacy degree k defining the size of the association groups into which each actual association protected by a confidentiality constraint must be subsumed. We also define properties that the grouping of tuples has to satisfy to guarantee a given privacy degree k of the associations while maximizing the information released (i.e., minimizing the size of the association groups) and respecting all the given confidentiality constraints.

Minimal Fragmentation

Given a relation, a set of confidentiality constraints, and a set of utility requirements, the challenge is to determine a correct fragmentation, that is, a fragmentation that is safe with respect to the constraints and satisfies the utility requirements. Correctness is formally defined as follows.

Correctness. Given a relation schema $S(a_1, ..., a_n)$, a set C of confidentiality constraints over S , and a set V of utility requirements over S , a fragmentation F of S is correct *wrt* C and V iff: i) no individual fragment $F \in F$ contains all the attributes involved in a confidentiality constraint (i.e., $\forall F \in F, \forall c \in C : c \nsubseteq F$); ii) fragments are disjoint (i.e., $\forall F_i, F_j \in F, i \neq j : F_i \cap F_j = \phi$); iii) $\forall v \in V, \exists F \in F : F$ can imply v.

Also, we aim at a minimal fragmentation, that is, a fragmentation that minimizes the number of fragments. Indeed, avoiding splitting attributes when not needed for satisfying the constraints is convenient, as it maximizes the utility over the data. In fact, maintaining attributes together in a fragment releases not only their values but their association, which, if not protected (directly or indirectly) by confidentiality constraints, can be safely released. Our problem is then formally defined as follows.

Min-CF. Given a relation schema $S(a_1, ..., a_n)$, a set C of confidentiality constraints over S, and a set V of utility requirements over S , determine (if it exists) a fragmentation F such that:

1. F is a correct fragmentation; 2. $\nexists F^*$, F^* is correct and the number of fragments of F^* is less than the number of fragments of F.

In the Task 1, we investigate data publishing methods based on: (i) privacy risk with fragmentation and confidentiality constraints and (ii) the Min-CF problem, identify the optimum solutions through correct and minimal fragmentations.

Task 2: Loose associations

This task will develop a fragment-involved publishing framework (*FIPF*). This framework includes a loose association model, supports efficient data sharing, utility and access query through fragmentations and illustrates how data owners interact with the system to manage fragmentations with both confidential constraints and utility requirements. Approaches in the access query can be applied for fragment-involved management in data publishing environments. However, no such access model exists and its development is an important challenge.

While fragments, by definition, cannot be joined (as this would imply a violation of confidentiality constraints), publishing a loose association among their tuples (sub-tuples of the original relation) can release some information on the association existing in the original relation, provided a given privacy degree of the association is respected. Intuitively, the loose associations hide tuples participating in the associations in groups and provide information on the associations only at the group level. Loose associations, while not impacting privacy (as dictated by the privacy degree), provide enriched utility of the published data, supporting, for example, aggregate queries and data mining.

We start by defining grouping over fragment instances. Since the size of the groups into which tuples in fragments are clustered impacts the association problem, we characterize a grouping with an index denoting the lower bound on the size that groups may have.

k-Grouping. Given a fragment instance f_i and a set GID_i of group identifiers, a k-grouping over f_i is a surjective function $G_i : f_i \rightarrow GID_i$ such that $\forall g_i \in GID_i : \|G_i^{-1}(g_i)\| >= k$.

A k-grouping function associates with each tuple in a fragment a group identifier in such a way that each group has at least k tuples mapping to it. A k-grouping is minimal if it minimizes the size of the groups, provided that k is respected, or, equivalently, it maximizes the number of groups into which tuples are mapped. Formally, a k-grouping over a fragment f_i is minimal if the cardinality of the image of G_i over f_i is equal to the floor of the ratio between the cardinality of f_i and k, that is, $\|GID_i\| = [\|f_i\|/k]$.

Intuitively, an association is k-loose iff for each real association existing in the original relation it releases at least k possible distinct associations. Figure 3 illustrates an example of loose association induced by the (F_l, F_r)-grouping in Figure 2.

In the following we consider the problem of grouping two fragments f_l and f_r, and we will refer to a (k_l, k_r)-grouping as a single term to denote the two components: a k_l- grouping over f_l and a k_r-grouping over f_r. A (k_l, k_r)- grouping is said to be minimal if both its grouping components are minimal.

This project will develop database systems for *FIPF* to support k-grouping, loose association and analyse the response time impact if data utility requirements are changed.

Task 3: Queries and data utility with loose associations

The reason for publishing group associations among fragments, representing vertical views over the original data, is to provide some (not precise) information on the associations among the tuples in the fragments while ensuring not to expose the sensitive associations defined among their attributes (for which the degree of uncertainty k should be maintained). Group associations then increase the utility of the data released for queries involving different fragments. However, given a set of fragments, different group associations might be defined satisfying a given degree k of looseness to be provided. Two different issues must be properly addressed in the construction of group associations: one is how to select the size k_i of the grouping of each fragment f_i such that the product of any two k_i is equal to or greater than k, and the other is how to group tuples within the fragments so to maximize utility.

With respect to the first issue of sizing the groups, different possible values of the different k_i can satisfy the degree k of protection. For instance, for a group association between two fragments, we can use $(k, 1)$, $([\sqrt{K}], [\sqrt{k}]$, and $(1, k)$. This task investigates the best utility that can be achieved in the case of multiple fragments.

With respect to the issue of grouping within a fragment, we first note that queries that involve a single fragment (i.e., all the attributes in the query belong to the same fragment) are not affected by fragmentation as they can be answered exactly by querying the fragment. For instance, with respect to the fragments in Figure 1, the following query involves attributes that belong to fragment F_r only.

q =SELECT AVG(Income) FROM PATIENTS GROUP BY Edu

Hence, the execution of the query over fragment F_r returns exactly the same result as its execution over the original relation PATIENTS in Figure 1(a). Queries involve two or more fragments, on which a group association is to be defined, with the goal of determining how to group the tuples in fragments so that the induced group associations maximize query utility.

The execution of queries over group associations brings in, together with the real tuples on which the query should be executed, all the tuples in their groups and the uncertainty (by definition) about which sub-tuples in a fragment are associated with which sub-tuples in other fragments. Our observation is therefore that groups within fragments should be formed so to contain as many tuples as possible that are similar for the attributes involved in the queries (have close values for continuous attributes). The reasoning behind this is that, although the query is evaluated on a possibly larger number of tuples included in the returned groups, such tuples (assuming similar values) maintain the query result within a reasonable error, thus providing utility of the response. The more the attributes involved in the query on which such an observation has been taken into account in the grouping, the better the utility provided by the group association for the query. In fact, similarity of values within groups (even when ensuring diversity of the values) might provide limited uncertainty of values within a group. We therefore expect that not all the attributes involved in confidentiality constraints should be taken into account in this process.

Let us see now an example of queries over our fragmentation involving attributes such that none, some or all of them have been subject to the observation above in the grouping. For example, a query q asks the average income amount for the different education levels recorded; this involves none of the attributes whose similarity has been considered in the grouping.

q = SELECT Edu, AVG(Income) FROM PATIENTS GROUP BY Edu

The task will analyse the utility impact of associations for this kind of queries.

Task 4: System design

We will design a prototype of the fragment-based framework for a public data set from CENSUS IPUMS (http://www.ipums.org) to demonstrate the feasibility of the proposed k-grouping methods and to provide loose association approaches for confidential constraints and utility requirements. The prototype will be a web-based application supporting secure sensitive information in data publishing systems. The web server and the application server are responsible for authenticating user identification information and providing confidential data transmission through secure socket layer connection. In this task, we study the design issues in general. We also plan to demonstrate the loose association model, utility requirements update and confidential protection in data publishing environments.

The working environment for data publishing in outsourced environments is highly distributed; customers' information must be protected from data access requests and the service provider as well as the systems that process the detection and prevention of clients' information. We will describe how fragmentations and loose associations are processed in the framework.

Task 4 is the preparatory work for an implementation of the results of this project to demonstrate the project's theoretical approach and contribution to new knowledge.

Task 5: Fragmentation approach evaluation

The main goals of the experimental evaluation for fragmentation are to investigate the overheads incurred by k-grouping and to measure the performance results of the proposed optimal fragmentation algorithms and loose association models.

We will study the overhead at the server side in the fragmentation task over the CENSUS data set. Instead of measuring performance in run time, we measure the increase in the number of attributes and constraints and utility requirements. This will test the scalability and the ability of the approach to manage large database schemas and complex privacy and utility requirements.

We plan to evaluate the precision of the queries executed on the fragments when a k-loose association is published, and how the error in the evaluation of queries evolves with the increase in the value of k in the approach.

Task 6: System implementation

We will implement the new techniques developed in Tasks 1-3 into a comprehensive prototype system for data publishing to satisfy confidential constraints and data utility requirements in outsourcing environments. We will develop graphical user interfaces (*GUI*) for both data owners and users to handle confidential constraints and utility requirements through access queries. Scalability will ensure the success of the presented techniques. We will interrogate a real-world data set to select the constraints occurring over a period of time, and create different data sets with varying numbers of utility requirements.

Task 6 focuses on a proof implementation to demonstrate the feasibility of the proposed loose association approaches and intelligent query algorithms, and provides efficient query protocols for users while protecting sensitive information in outsourced data sets.

FEASIBILITY

The proposed project represents a collective effort of *CIs* Wang and Zhang at *VU* and *CI* Cao at *La Trobe University*. The *CIs* have long term research collaborations and published 23 high quality research papers. The *CIs* have complementary expertise in data publishing and data privacy, intelligent algorithm design and evaluation, and query processing and optimization. All these skills are required to perform the challenging tasks of the proposed project. Both *CIs* Wang and Zhang are world-leading expert in data mining, privacy modelling and design and has strength in access modelling and query processing; *CI* Cao is an active researcher in outsourced data and expert in algorithm evaluation, query processing and data security. The formation of such a team and training of junior researchers will greatly increase the critical mass of internationally competitive research capabilities in this important area.

Time Feasibility. Our previous work on privacy preserving data publishing is the foundation for our proposed fragmentation and access query approaches. We will develop optimal fragmentation algorithms to achieve k-grouping rules by extending our previous work, which will help increase our efficiency in completing Tasks 1, 2 and 3. We can then focus on loose association models and algorithms of confidential constraints and data utility that will ensure that we complete Tasks 1-3 in two years. Then, we will evaluate the loose association ideas and data utility rules using a real data set in the final year. Our previous work and well-planned tasks ensure that the project can be completed in three years. A detailed timeline is shown in Table 3.

Period	Tasks
Jan 2018 - Dec 2018	Task 1 (CIs Wang, Zhang, Cao, RA, PhDs), Task 2 (CIs Wang, Zhang, RA, PhDs)
Jan 2019 - Dec 2019	Task 3 (CI Wang, Zhang, Cao and RA, PhD1), Task 4 (CIs Wang, Zhang, Cao, PhD2, RA)
Jan 2020 - Dec 2020	Task 5 (CIs Wang, Zhang, Cao, PhD1, RA), Task 6 (CIs Wang, Cao, RA, PhD1, PhD2)

Table 3. Timeline of the project

Method and Experiment Feasibility. This project will employ *PUMAs* data sets from the American Community Survey (*https://www.census.gov/geo/reference/puma.html*). The survey involves a 3-in-100 random sample of the population of the United States and the data set contains all households and persons from the 1% ACS samples for 2010, 2011, and 2012, identifiable by year. The data sets with *PUMAs* contain real-world information for at least 100k persons that will support the approaches developed in the project.

We will use the above data for studying fragmentation and data utility. Implementing our proposed methods and their demonstrations using the above data are feasible; we can build benchmarks to evaluate the accuracy of the loose associations and efficiency of the intelligent algorithms. Also, our collaborative industry network including linkage partners will ensure that validation of our methods in these applications will have broad applicability.

Commitment to research: *VU* has a world-class research environment in the fields of Artificial Intelligence and Image Processing and Distributed Computing, according to the Excellence in Research for Australia (ERA) 2015 evaluation; these are the two major research fields of the *CAI*. *VU*'s Research Plan 2012-2016 aims to considerably increase its research activity and standing among Australian universities by focusing on applied and translational research. *VU*'s goal is to create a vibrant research culture and be known for the international standard of its research. La Trobe University has defined research strategy that encourages innovation, specialisation and collaboration to address important challenges in the 21st century.

Facilities, infrastructure, equipment: *VU* has a long tradition as a research-active environment with regular staff and postgraduate seminars, funding for conference travel and excellent research in frastructure. Both *VU* and La Trobe University have excellent facilities for applied informatics research, due to their support of Applied Informatics as one of its strategic research areas.

Cash contribution: *VU* will make cash contribute $113K to the project. 1) *VU* will contribute cash $37,550 pa for 3 years for an PhD stipend ($26,300) and international conferences ($11,250).

Adequacy of Research Environment

The *CAI* focuses on data mining, privacy preserving and e-research with applications in outsourcing environments and has an international reputation in these areas. *CAI*'s research has been funded by 11 ARC grants since 2003. *CAI* has worked successfully with government and industry organisations to apply its core techniques in data mining and privacy management in e-Health and e-Environment fields. The *CIs* have established close collaborative relationships through frequent academic visits and cooperative research projects, and have co-authored many high-quality publications, for example:

1. H. Wang, J Cao and Y Zhang. *A flexible payment scheme and its role-based access control*. IEEE Transactions on Knowledge and Data Engineering, 17 (3), 425-436, 2005. (Q1, Citation 74)

2. J. Huang, M. Peng, H. Wang, J. Cao, G. Wang, X. Zhang. *A Probabilistic Method for Emerging Topic Tracking in Microblog Stream*, Accepted by World Wide Web on Feb, 2016. (Q1)

The existing active collaboration and complementary expertise of the research team put them in an excellent position to succeed in the proposed project.

BENEFIT

National benefit The national benefits of this project are manyfold. First, private information protection and data utility in outsourced environments is a core technology needed in many intelligent systems with applications in data processing, fraud detection and counterterrorism. Second, improved techniques for fragmentation, data protection, standardising and sharing advance and facilitate the processing and analysis of information within many organi-

sations. Third, advanced data access techniques for privacy protection have wide application in crime and fraud detection, which will form an important part of future counterterrorism prevention and monitoring systems. This project thus contributes to the Challenge of **Highly-secure and resilient communications and data acquisition, storage, retention and analysis for government, defence, business, transport systems, emergency and health services** within the Cybersecurity Research Priorities as it will ultimately lead to better utilisation and analysis of the information available in many organisations.

COMMUNICATION OF RESULTS

The project results will be communicated through several channels.

a. Refereed international journals (e.g. IEEE Transactions on Pattern Analysis and Machine Intelligence, Pattern Recognition, Information Systems, IEEE TKDE) and through presentations to international conferences (ICDM, ICDE, AAAI, etc). Publication of technical reports, white papers and patent applications is anticipated.

b. A prototype system. We will incorporate the newly developed framework, data fragmentation algorithms and techniques into prototype software which will be made available on the Internet.

c. Seminars at *VU* and La Trobe University, and engagement with industry, government and community.

d. Two PhD thesis will be produced during the project. Further, selected materials from the project will be incorporated into postgraduate level teaching at *VU* and La Trobe University.

MANAGEMENT OF DATA

Both Victoria University and La Trobe University support the management of research data through the training of staff and students and by providing resources for securely maintaining data and materials. The Universities provide enterprise-grade, secure research storage and backup for safe storage during research and for long-term retention. Physical records and material are retained within the universities' colleges. The CIs will ensure the responsible management of data, materials and records during the project, and that after the research concludes the project data is retained in a durable format and can be appropriately accessed. Data and materials will be made available for reuse by postgraduate students at Victoria University and La Trobe University.

VU registers collections/data for reuse in Research Data Australia, the national *ANDS* (*https:www.ands.org.au*) registry. Victoria University also hosts *https:data.vu.edu.au*, an institutional repository that makes appropriate digital collections more widely available for reuse.

REFERENCES

1. P. Berman and S. Raskhodnikova. Approximation Algorithms for Min-Max Generalization Problems. *ACM Transactions on Algorithms*, 11(1), doi¿10.1145/2636920, 2014.
2. J. Biskup. Selected results and related issues of confidentiality-preserving controlled interaction execution. *Proceedings of the International Symposium on Foundations of Information and Knowledge Systems*, pp: 211-234, 2016.
3. J. Kohler, K. Jnemann, H. Hartenstein. Confidential database-as-a-service approaches: taxonomy and survey. *Journal of Cloud Computing*, 4(1): 1-14, 2015.
4. S. De Capitani di Vimercati, et al. Loose associations to increase utility in data publishing1. *Journal of Computer Security*, 23(1): 59-88, 2015.
5. S. De Capitani di Vimercati, et al. Fragmentation in presence of data dependencies. *IEEE TDSC*, 11 (6): 510 - 523, 2014.
6. C. Dwork. Differential privacy. *Proc. of ICALP*, pp: 1-12, 2006.
7. C. Gentry and S. Halevi. Implementing Gentrys fully-homomorphic encryption scheme. *Proc. of EUROCRYPT*, pp: 129-148, 2011.
8. J. Huang, M. Peng, H. Wang, J. Cao, G. Wang, X. Zhang. A probabilistic method for emerging topic tracking in Microblog stream, *World Wide Web*, 20(2): 325350, 2017.
9. E. Kabir, A. Mahmood, H. Wang, and A. Mustafa. Microaggregation sorting framework for k-anonymity statistical disclosure control in cloud computing. *IEEE Transactions on Cloud Computing*, DOI:10.1109/TCC.2015.2469649, 2015.
10. K. Lin and M. Chen. On the design and analysis of the privacy-preserving SVM classifier. *IEEE Transactions on Knowledge and Data Engineering*, 23(11): 1704-1717, 2011.
11. R. Jhawar, V. Piuri, and P. Samarati. Supporting security requirements for resource management in cloud computing. *Proc. of CSE 2012*, pp: 170-177, 2012.
12. T. Jiang, X. Chen, and J. Ma. Public integrity auditing for shared dynamic cloud data with group user revocation. *IEEE Transactions on Computers*, 65(8): 2363-2373, 2016.
13. P. Samarati. Data security and privacy in the cloud. *Information Security Practice and Experience*, pp: 28-41, 2014.
14. Y. Sun, et al. Attacks and countermeasures in the internet of vehicles. *Annals of Telecommunications*, pp: 1-13, 2016.
15. J. Tang, et al. Ensuring security and privacy preservation for cloud data services. *ACM Computing Surveys (CSUR)*, 49(1), doi¿10.1145/2906153, 2016.
16. J. Yu, G. Wang, Y. Mu, and Gao, W. An efficient generic framework for three-factor authentication with provably secure instantiation. *IEEE Transactions on Information Forensics and Security*, 9(12): 2302-2313, 2014.
17. H. Wang, Y. Zhang, J. Cao. Effective collaboration with information sharing in virtual universities, *IEEE Transactions on Knowledge and Data Engineering*, 21(6): 840-853, 2009.

C2. Medical Research

(Does this Project contain content which requires a statement to demonstrate that it conforms with the eligible research requirements set out in the ARC Medical Research Policy located on the ARC website?)

No

C3. Medical Research Statement

(If applicable, in no more than 750 characters (approx.100 words), justify why this Project complies with the eligible research requirements set out in the ARC Medical Research Policy located on the ARC website.)

C4. Please list the objectives of your proposed Project.

(Please list each objective separately by clicking 'add answer' to add the next objective. You may enter 500 characters (approximately 70 words) per objective.

This information will be used for future reporting purposes if this Proposal is funded.

(This question must be answered))

Objective

develop a fragmentation-based framework to protect confidentiality and satisfy utility requirements when outsourcing data to external storage.

Objective

design an innovative logical language for the utility of loose associations in terms of providing better response to queries is proposed as the basic technique within this framework

Objective

invent access query algorithms by using the original data sets and newly generated data sets.

Objective

implement a prototype set of loose association rules over synthetic and real data sets aimed at evaluating the efficiency, efficacy and scalability of the developed algorithm.

Part D - Project Cost (DP180103563)

D1. What is the proposed budget for your Project?

(Please refer to the Instructions to Applicants for detailed instructions.)

Total requested budget: $443,451

Year 1

Description	ARC	Admin Org		Other Eligible Org	
	Cash	Cash	In-kind	Cash	In-kind
Total	147,817	37,550	84,906		31,806
Personnel	136,567	26,300	84,906		31,806
CI(Wang)Level E@0.2FTE + 30% on-cost			42,453		
CI(Zhang)Level E@0.2FTE + 30% on-cost			42,453		
CI(Cao)Level C@0.2FTE + 30% on-cost					31,806
Research Assistant (Level B Step 2) + 30%	110,267				
Higher Degree by Research stipend		26,300			
Higher Degree by Research stipend	26,300				
Travel	11,250	11,250			
4 domestic and 3 overseas conferences (4*2000 + 3* 5500)	11,250	11,250			

Year 2

Description	ARC	Admin Org		Other Eligible Org	
	Cash	Cash	In-kind	Cash	In-kind
Total	147,817	37,550	87,876		32,442
Personnel	136,567	26,300	87,876		32,442
CI(Wang)Level E@0.2FTE + 30% on-cost			43,938		
CI(Zhang)Level E@0.2FTE + 30% on-cost			43,938		
CI(Cao)Level C@0.2FTE + 30% on-cost					32,442
Research Assistant (Level B Step 2) + 30%	110,267				
Higher Degree by Research stipend		26,300			
Higher Degree by Research stipend	26,300				
Travel	11,250	11,250			
4 domestic and 3 overseas conferences (4*2000 + 3* 5500)	11,250	11,250			

Year 3

Description	ARC	Admin Org		Other Eligible Org	
	Cash	Cash	In-kind	Cash	In-kind
Total	147,817	37,550	91,832		33,091
Personnel	136,567	26,300	91,832		33,091

CI(Wang)Level E@0.2FTE + 30% on-cost			45,916		
CI(Zhang)Level E@0.2FTE + 30% on-cost			45,916		
CI(Cao)Level C@0.2FTE + 30% on-cost					33,091
Research Assistant (Level B Step 2) + 30%	110,267				
Higher Degree by Research stipend		26,300			
Higher Degree by Research stipend	26,300				
Travel	11,250	11,250			
4 domestic and 3 overseas conferences (4*2000 + 3* 5500)	11,250	11,250			

Other Eligible Organisation

Organisation	Year 1		Year 2		Year 3	
	Cash	In-kind	Cash	In-kind	Cash	In-kind
La Trobe University		31,806		32,442		33,091
Total		31,806		32,442		33,091
Committed Total		31,806		32,442		33,091

Part E - Budget Justification (DP180103563)

E1. Justification of funding requested from the ARC

(In no more than five A4 pages and within the required format fully justify, in terms of need and cost, each budget item requested from the ARC. Use the same headings as in the Description column in the budget at Part D of this Proposal.)

Uploaded PDF file follows on next page.

Justification of funding requested from the ARC

Funding is requested for **(1) Personnel:** One Research Assistant and one PhD Stipends; **(2) Travel:** three domestic conference travel and three international conference travels.

PhD student: One PhD student funded by ARC will work on data outsourcing model with privacy risk and confidential constraints and minimal fragmentation methods, design SQL query model, the relationship between fragmentation and utility approaches (Tasks 1 and 3). Additionally, the student is required to work on the demonstration of the access queries and validations in real datasets (Task 5). An ARC scholarship rate (total of $ 26,300 per annum for three years) is requested.

Research Assistant (RA): One full time research associate (RA) at level B step 2 with the skill base typical of PhD in computer science is requested to undertake creating, investigating and comparing various access queries and k-grouping methods, and also implementation and evaluation the designed models and approaches in a real dataset system (Tasks 1, 2, 3, 5 and 6). Task 5 and 6 involve a large amount of implementations and experiments, and hence an experimental RA is required. The cost for hiring a Research Assistant @ Level B step 2 +30% on-cost: $110,267, $113,575, $116,982 for the first, second and third year respectively.

The RA will also be involved in investigating association rule analysis and interactive analysis with pattern mining and data privacy methods, and developing access query to detect the efficiency of the fragmentation algorithms (Tasks 1 and 2) in Year 1, and comparing and evaluation various access query and utility methods (Task 3) in Year 2, and implementation of data outsourcing models and system evaluation for a CENSUS dataset in outsourced environments (Tasks 5 and 6) in Year 3.

2. Travel: request $11,250 pa. for three years from ARC
We target on AAAI, VLDB, ICDE, ICDM and SIGMOD/PODS to present the project progress and challenges with experts in conferences. Participation at conferences, to communicate results and seek new information and views, is an essential part of research. For the six participants we plan a sum of $24,500 per year to support participation in a total of three Australian (RA, PhDs) and three overseas conferences (CIs Wang, Zhang, Cao).

One domestic conference:
Registration fee ($800) +Domestic flight ($400) + 4 nights' accommodation (4*$200)
= 2,000
Subtotal for four domestic conferences: $2000*3 = $6,000

One overseas conference:
Registration fee ($1000), flight from Melbourne to US or Europe ($3000), 5 nights' accommodation (5*$300)
Subtotal for three overseas conferences: $5500*3 = $16,500

Total: $22,500 (per year)

These costs are split evenly between the universities and the ARC - each $11,250 per annum.

E2. Details of non-ARC contributions

(In no more than two A4 pages and within the required format, provide an explanation of how non-ARC contributions will support the proposed Project. Use the same headings as in the Description column in the budget at Part D of this Proposal.)

Uploaded PDF file follows on next page.

Details of Non-ARC contributions

Victoria University will make cash contribution $113K in total to the project.

The structure of the contributions from both Victoria University and La Trobe University to the project:

1. Contribute 0.2 FTE of salary ($42,453 (yr1), $43,938 (yr2), $45,916 (yr3)) from VU for Prof. Wang;

2. Contribute 0.2 FTE of salary ($42,453 (yr1), $43,938 (yr2), $45,916 (yr3)) from VU for Prof. Zhang;

3. Contribute 0.2 FTE of salary ($31,806 (yr1), $32,442 (yr2), $33091 (yr3)) from La Trobe University for Dr Cao;

4. Cash contribution $113K from VU to support a PhD stipend ($79K) and half of the estimated conference travel expenses for the participants in the project ($11,250*3).

   The second PhD student funded by Victoria University will be co-supervised by the three CIs. The student will involve the analysis and development of the fragment-based intelligent algorithms (Task 2), design a prototype of the fragment-based framework for a public data set to demonstrate the feasibility of the proposed k-grouping methods (Task 4); develop graphical user interfaces for both data owners and users to handle confidential constraints and utility requirements through access queries (Task 6).

Part F - Personnel and ROPE (Prof Hua Wang)

F1. Personal Details

(To update personal details, click the 'Manage Personal Details' link below. Note this will open a new browser tab. When returning to the form ensure you 'Refresh' the page to capture the changes made to your profile.)

Participation Type

Chief Investigator

Title

Prof

First Name

Hua

Second Name

Family Name

Wang

F2. Will you be residing predominately in Australia for the duration of the Project?

(This is a 'Yes' or 'No' question. Indicate whether you will be residing predominantly in Australia for the duration of the Project, taking into account any international travel. If you are applying as a CI and you answer 'No' to this question you will be prompted to contact your Research Office to check your eligibility.)

Yes

F3. Qualifications

(To update any qualifications, click on the 'Manage Qualifications' link below. Note this will open a new browser tab. When returning to the form ensure you 'Refresh' the page to capture the changes made to your profile.)

Conferral Date	AQF Level	Degree/Award Title	Discipline/Field	Awarding Organisation	Country of Award
15/04/2004	Doctoral Degree	PhD	Information technology	University of Southern Queensland	Australia

F4. Are you currently undertaking a Higher Degree by Research which will be conferred after 1 January 2018?

(This is a 'Yes' or 'No' question. If you are applying as a CI and your answer is 'Yes' to this question you will be prompted to contact your Research Office.)

No

F5. Research Opportunity and Performance Evidence (ROPE) – Current and previous appointment(s) / position(s) - during the past 10 years

(To update any details in this table, click on the 'Manage Employment Details' link below. Note this will open a new browser tab. When returning to the form ensure you 'Refresh' the page to capture the changes made to your profile. Click on the information icon above and refer to the Instructions to Applicants for more information.)

Description	Department	Contract	Employment	Start Date	End Date	Organisation

		Type	Type			
Professor	College of Engineering and Science	Permanent	Full Time	27/02/2014		Victoria University
Professor	Maths & Computing	Permanent	Full Time	01/01/2011	26/02/2014	University of Southern Queensland
Associate Professor	Maths & Computing Department	Permanent	Full Time	01/01/2009	31/12/2010	University of Southern Queensland
Senior Lecturer	Maths & Computing Department	Permanent	Full Time	01/01/2007	31/12/2008	University of Southern Queensland

F6. Employment Details as at Commencement date of Project

(This question will be used to determine your eligibility. Confirm your employment status at all organisations that you will be associated with as at the Commencement Date for the Project (1 January 2018). Enter the relevant appointment type and Full-Time Equivalent (FTE) for each organisation. Click on the information icon for further information.)

Org name	Is this an Eligible Organisation?	Please choose your appointment type for this organisation.	Please enter your FTE for this Organisation
Victoria University	Yes	Employee	1

F7. Further Details Regarding Partner Investigator Status - Do you hold a remunerated appointment at an Eligible Organisation?

(NOTE: this question is mandatory ONLY FOR PIs WHO:
• at F2 confirmed that they will reside predominantly in Australia for the duration of the proposed Project; AND
• at F4 confirmed that they are not currently undertaking a Higher Degree by Research which will be conferred after 1 January 2018; AND
• at F6 indicated that they would hold either:
- an appointment at an Eligible Organisation equal or greater than 0.2FTE; OR
- Emeritus appointment at an Eligible Organisation

This is a 'Yes' or 'No' question. If you select 'Yes', you will be further prompted to justify your participation on this Proposal as a PI with reference to sections F7.2 and F7.3 of the Funding Rules. Click on the information icon for further information.)

Do you hold a remunerated appointment at an Eligible Organisation?

Justification of PI status

F8. Relevant Organisation for this Proposal

(Enter the Organisation that is relevant to your participation on this proposal, and that you will be associated with as at the Commencement Date for the Project (1 January 2018). The 'relevant organisation' is the primary organisation that will be supporting your involvement in this Project if it is funded. Note that the Organisation must be listed in F6 for this question to validate.)

Relevant Organisation

Victoria University

F9. What is your time commitment to this Project?

(Enter your time commitment to this Project as a Full-Time Equivalent (FTE). Note that a FTE of 1.0 represents a full-time commitment (i.e. 5 days per week).)

0.2

F10. Are you applying for Teaching Relief?

(This is a 'Yes' or 'No' question.

• If you answer 'Yes' to this question a budget line will be automatically populated for the Teaching Relief in the budget table in Form Part D: Project Cost. This will allow you to enter the funding amount requested in the relevant year/s. To remove the Teaching Relief from the budget table you must return to this question and answer 'No'.

• Note: CIs may request funding for teaching relief or other duties in order to maximise the opportunity for the CI to conduct research. This question is only relevant for CIs and will not be activated for PIs. If the answer to this question is 'yes', the budget table in 'Project Costs' will automatically update with a line to add costs for Teaching Relief.)

No

F11. Are you applying for a Discovery International Award?

(This is a 'Yes' or 'No' question. If you answer 'Yes' to this question a budget line will be automatically populated for the DIA in the budget table in Form Part D: Project Cost. This will allow you to enter the funding amount requested in the relevant year/s. To remove the DIA from the budget table you must return to this question and answer 'No'. Click on the information icon for further information.)

No

F12. Research Opportunity and Performance Evidence (ROPE) - Details of your career and opportunities for research

(Write a maximum of 5250 characters (approximately 750 words). Please detail your career and opportunities.)

Hua Wang received his PhD degree in Computer Science at the University of Southern Queensland (USQ) in 2004 and was an Early Career Researcher until 2009. He has been employed at USQ from 2003 as a lecturer. He joined the Centre of Science of Biology (CSBi) as a Bioinformatics division leader from 2007, involving research in database systems, data mining, privacy, access control and e-health information systems. Hua has supervised seven PhD students in the past ten years, and is currently supervising seven PhD students (five principal and two associate).

Hua became Professor at Victoria University in 2014. Hua has had a mixed duty of teaching, research and administration/service allocated as 10%, 80%, and 10% time, respectively. Hua is a full-time academic (research and teaching) and has had no career interruption. Hua engages collaboratively with several high-profile international research groups. In 2006, hua spent six months to work with Chair Professor Xiaohua Jia at Hong Kong City University, and he won an Japan Society Promotion Scholarship (JSPS) in 2008 and 2015. Hua joined Professor Jeffery Yu's database group at Hong Kong Chinese University in 2009, and worked for Cloud Computing Centre at the Stavanger University in Norway in 2011. All these international collaborations enrich Hua's science view on challenges and problem analysis.

Additionally, Hua has worked closely with both CIs Zhang and Cao in data mining and database systems research. Many scholar papers have been published as outcomes of the collaborative work. For example:

• Supriya, Siuly, H. Wang, J. Cao and Y. Zhang. Weighted visibility graph with complex network features in the detection of epilepsy. IEEE Access, Vol. 4, pp: 6554 - 6566, 2016. (Q1; IF=1.270)

• H. Wang, J. Cao, Y. Zhang. A flexible payment scheme and its role based access control. IEEE Transactions on Knowledge and Data Engineering (TKDE). 17(3): 425-436, 2005 (74 citations).

• *H. Wang, Y. Zhang, J. Cao. Effective collaboration with information sharing in virtual universities. IEEE Transactions on Knowledge and Data Engineering. 21(6): 840-853, 2009. Outcomes from ARC DP0663414 (40 citations).

In the past years, Hua has published 166 scholar research papers including 64 refereed journal articles and 95 refereed conference papers in data mining, privacy, data anonymisation, and text classifications. He is co-editor-in-chief for the peer reviewed journal ICST Transaction on Scalable Information Systems. His research has attracted over AU$1M funding since 2006 from Australia and Japan, including three ARC Discovery grants and one ARC Linkage project.

Hua has two research mentors: (1) Professor Yanchun Zhang who was his PhD supervisor ten years ago and he is continuously the mentor of Hua. Professor Zhang has abundant experience in mentoring young researchers; and (2) Professor Elisa Bertino at Purdue University who has supervised Hua with collaborative papers and assessed Hua's previous ARC DP projects.

Hua is committed to applied research outcomes that resolve real-world problems. His research includes one ARC Linkage grant related to health environment and three ARC Discovery grants related to data mining and privacy:

• LP150100673, Privacy Preserving Data Sharing in Electronic Health Environment, (Y. Zhang, H. Wang et al, $295,000 + $150,000, 2016-2018)
• DP0988465, Limiting disclosure of private information in relational database systems (H. Wang, $245000, 2009-2011).
• DP0774450, Privacy preserving data sharing in data mining environments (J. Li and H. Wang, $186000, 2007-2009).
• DP0663414, Protect information sharing within distributed collaborative environment (H. Wang, $98000, 2006-2008).

These ARC projects provide Hua with the necessary physical and human resources to collaborate effectively with industry partners and develop his own research direction. These ARC projects also demonstrate that Hua's research interests in many subfields in database and privacy protection are valuable and feasible. Hua is a research active person and was the chair of early career research group at USQ. Furthermore, he is the Co-Editor in Chief of the international journal "ICST Transaction on Scalable Information Systems". This experience broadened his scientific views and enriched his research skills.

VU has a world class research environment in the fields of Artificial Intelligence and Image Processing and Distributed Computing, according to the Excellence in Research for Australia (ERA) 2012 evaluation; these are the two major research fields of the CAI. VU's Research Plan 2012-2016 aims to considerably increase its research activity. VU's goal is to create a vibrant research culture and be known for the international standard of its research. VU will make $113K/year cash contribution to the project.

In summary, Hua's strong capability of achieving high quality research outcomes in the past years and his extensive research program on data mining, architecture intelligence, and privacy preserving means that this proposed project is highly feasible. Hua's successful independent conduct of complex projects and PhD student supervision demonstrates his leadership abilities.

F13. Research Opportunity and Performance Evidence (ROPE) - Significant research outputs and ARC grants

(Upload a PDF of no more than four A4 pages with a list of all research outputs, such as journal articles and refereed conference papers, book and book chapters. Use asterisks to identify research outputs relevant to this Proposal. Click on the information icon or refer to the Instructions to Applicants for the required content and formatting.)

Uploaded PDF file follows on next page.

Recent significant publications and ARC grants: Hua Wang

Hua has published 166 refereed scholar papers, including 39 Q1 Journal papers in data mining, database, privacy preserving and Web services. Representative publications are on *ACM Transactions on Internet Technology, IEEE Transactions on Dependable and Secure Computing, Computational and Mathematical Methods in Medicine, Data Mining and Knowledge Discovery, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Automation Science and Engineering and World Wide Web;* and Proceedings of *AAAI, CIKM and ICDM and PAKDD*. His work has been highly cited in recent years with H-index: 22 and i10-index: 44.

REFERRED JOURNAL ARTICLES:

1. Supriya, Siuly, **H. Wang**, J. Cao and Y. Zhang. *Weighted visibility graph with complex network features in the detection of epilepsy*. IEEE Access, DOI: 10.1109/ACCESS.2016.2612242, 2016. (Q1; IF=1.270)
2. Y. Wang; Y. Shen; **H. Wang**; J. Cao; X. Jiang. *MtMR: Ensuring MapReduce Computation Integrity with Merkle Tree-based Verifications*, IEEE Transactions on Big Data, no.99, pp: 1-16, 2016.
3. *J. Huang, M. Peng, **H. Wang**, J. Cao, G. Wang, X. Zhang. *A Probabilistic Method for Emerging Topic Tracking in Microblog Stream*, World Wide Web, 20(2): 325-350, 2017. (Q1, IF: 1.47)
4. D. Pandey, X. Yin, **H. Wang**, Y. Zhang. *Accurate Vessel Segmentation using Maximum Entropy incorporating Line Detection and Phase-Preserving Denoising*, Accepted by Computer Vision and Image Understanding on December, 2016. (Q1, IF: 2.134)
5. S. Siuly, **H. Wang** and Y. Zhang. *Detection of motor imagery EEG signals employing Naïve Bayes based learning process*. Measurement. Volume 86: 148–158, 2016 (Q2, IF: 1.48)
6. Y. Zhang, Y. Shen, **H. Wang**, J. Yong, X. Jiang, *On Secure Wireless Communications for IoT under Eavesdropper Collusion*, IEEE Transactions on Automation Science and Engineering, Volume: 13, Issue: 3, pp: 1281 – 1293, 2016 (Q1, IF: 2.43)
7. *J. Zhang, H. Li, X. Liu, Y. Luo, F. Chen, **H. Wang**, L. Chang. *On Efficient and Robust Anonymization for Privacy Protection on Massive Streaming Categorical Information*. IEEE Transactions on Dependable and Secure Computing, no.1, pp. 1, 01/2015; DOI: 10.1109/TDSC.2015.2483503 (Q1, IF: 1.35)
8. Y. Zhang, Y. Shen, **H. Wang**, Y. Zhang, X. Jiang, *On Secure Wireless Communications for Service Oriented Computing*, IEEE Transactions on Services Computing, , no. 1, pp. 1, doi:10.1109/TSC.2015.2478453 (Q1, IF: 3.05)
9. *J. Ma, L. Sun, **H. Wang**, Y. Zhang and U. Aickelin, *Supervised Anomaly Detection in Uncertain Sensor Data Streams*, ACM Transactions on Internet Technology (TOIT), 16, 1, Article 4 (January 2016), 20 pages. DOI= http://dx.doi.org/10.1145/2806890 (Q1, IF: 0.77)
10. S. Siuly, E. Kabir, **H. Wang**, and Y. Zhang, *Exploring Sampling in the Detection of Multicategory EEG Signals*, Computational and Mathematical Methods in Medicine, vol. 2015, Article ID 576437, 12 pages, 2015. doi:10.1155/2015/576437 (Q2, IF: 0.77)
11. L. Sun, J. Ma, **H. Wang**, and Y. Zhang, *Cloud Service Description Model: An Extension of USDL for Cloud Services*, IEEE Transactions on Services Computing. 2015, http://doi.ieeecomputersociety.org/10.1109/TSC.2015.2474386 (Q1, IF:3.05)
12. *E. Kabir, A. Mahmood, **H. Wang**, A. Mustafa, *Microaggregation Sorting Framework for K-Anonymity Statistical Disclosure Control in Cloud Computing*, IEEE Transactions on Cloud Computing, vol. PP, no.99, pp.1-1, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7208829&isnumber=6562694
13. **H. Wang**, L. Sun and E. Bertino. *Building access control policy model for privacy preserving and testing policy conflicting problems*. Journal of Computer and System Sciences. 80 (8): 1493–1503, 2014, (Q1, IF: 0.48)
14. Y. Qin, Q. Sheng, N. Falkner, S. Dustdar, **H. Wang**, A. Vasilakos. *When Things Matter: A Data-Centric View of the Internet of Things*. J. Network and Computer Applications 64: 137-153 (2016). (Q1, IF: 2.23)
15. ***H. Wang**, X. Yi, E. Bertino and L. Sun. *Protecting outsourced data in cloud computing through access management.* Concurrency Computat.: Pract. Exper.. doi: 10.1002/cpe.3286, 2014 (Q2, IF: 0.997)
16. *J. Zhang, X. Tao, **H. Wang**. *Outlier detection from large distributed databases*. World Wide Web 17(4): 539-568 (2014). (Q1, IF: 1.47)
17. X. Sun, **H. Wang** and J. Li, and Y. Zhang. *Satisfying Privacy Requirements Before Data Anonymization*. The Computer Journal, Vol. 55(4), pp. 422-437, 2012 (IF: 0.79, citations 11)
18. L. Sun, **H. Wang**. *A purpose-based access control in native XML databases*. Concurrency and Computation: Practice and Experience, Vol. 24(10), pp. 1154-1166, 2012 (Q2, IF: 1.0, citations 8)
19. *X. Sun, **H. Wang** and Y. Zhang. *On the identity anonymization of high-dimensional rating data*. Concurrency and Computation: Practice and Experience, Vol. 24(10), pp. 1108-1122, 2012 (Q2, IF: 0.997)
20. M. Kabir, **H. Wang**, E. Bertino: *A role-involved purpose-based access control model*. Information Systems Frontiers Vol. 14(3), pp. 809-822, 2012 (Q2, IF:1.08, citations 19)
21. *X. Sun, **H. Wang** and J. Li, and Y. Zhang. *An approximate microaggregation approach for microdata protection*. Expert Syst. Appl. Vol.39(2), pp. 2211-2219, 2012 (Q1, IF: 2.0, citations 6)

22. M. Li, X. Sun, **H. Wang**, Y. Zhang, J. Zhang. *Multi-level delegations with trust management in access control systems*. J. Intell. Inf. Syst. Vol. 39(3), pp. 611-626, 2012. (Q1, IF: 0.89)
23. M. Li, X. Sun, **H. Wang**, Y. Zhang, J. Zhang. *Privacy-aware access control with trust management in web service.* World Wide Web Vol. 14(4), pp. 407-430, 2011 (Q1, IF: 1.47, citations 21)
24. *X. Sun, **H. Wang**, J. Li, J. Pei. *Publishing anonymous survey rating data*. Data Min. Knowl. Discov. Vol. 23(3), pp. 379-406, 2011, (Q1, IF: 1.99, citations 12)
25. *X. Sun, **H. Wang** and J. Li, and Y. Zhang. *Injecting purpose and trust into data anonymisation*. Computers & Security, Vol. 30(5), pp. 332-345, 2011. (Q1, IF: 1.39, citations 20)
26. M. Kabir, **H. Wang** and E. Bertino. *Efficient Systematic Clustering for k-anonymization*, Acta Informatica, Vol. 48(1), pp. 51-66 (Q2, IF: 0.47, citations 11)
27. L. Sun, **H. Wang**. *Access control and authorization for protecting disseminative information in E-learning workflow*. Concurrency and Computation: Practice and Experience, 23(16): 2034-2042, 2011 *(Q2, IF:1.0)*
28. *X. Sun, **H. Wang** and J. Li, and Y. Zhang. *Injecting purpose and trust into data anonymisation*. Computers & Security, Vol. 30(5), pp. 332-345, 2011. (Q1, IF: 1.39, citations 20)
29. *X. Sun, **H. Wang**, J. Li: *Validating Privacy Requirements in Large Survey Rating Data*, Next Generation Data Technologies for Collective Computational Intelligence, pp. 445-46, 2011.
30. *X. Sun, M. Li and **H. Wang**. *A Family of Enhanced (L; \alpha)-Diversity Models For Privacy Preserving Data Publishing*. Future Generation Computer Systems. 27(3): 348-356, 2011. (Q1, IF: 2.79, citations 16)
31. L. Sun and **H. Wang**, *Access control and authorization for protecting disseminative information in E-learning workflow*, Concurrency and Computation: Practice and Experience, 23:2034–2042, 2011 (Q2, IF: 1.0)
32. L. Sun and **H. Wang**, *A purpose based access control in native XML databases*, Concurrency and Computation: Practice and Experience, Vol. 24(10), pp. 1154–1166, 2012 (Q2, IF: 1.0)
33. *X. Sun and **H. Wang**, *On the Identity Anonymization of High Dimensional Rating Data*, Concurrency and Computation: Practice and Experience, Vol. 24(10), pp. 1108-1122, 2012 (Q2, IF: 1.0)
34. *X. Sun, **H. Wang** and L. Sun. *Extended k-Anonymity Models Against Sensitive Attribute Disclosure*. Computer Communication, Vol. 34(4), pp. 526-535, Elsevier, 2011. (Q1, IF: 1.70, citations 27)
35. J. Zhang, Q. Gao, **H. Wang**, H. Wang: *Detecting anomalies from high-dimensional wireless network data streams: a case study*. Soft Computer, Vol. 15(6), pp. 1195-1215, 2011. (Q1, IF:1.27)
36. M.E., Kabir, **H. Wang** and E. Bertino, *A Conditional Role-involved Purpose-based Access Control Model*, Journal of Organizational Computing and Electronic Commerce, Vol. 21(1), pp. 71-91, 2011 (Q1, IF: 0.88)
37. L. Sun and **H. Wang**, *A Purpose Based Usage Access Control Model*, International Journal of Computer and Information Engineering, 4:1 2010, pp. 44-51.
38. M.E., Kabir, **H. Wang** and E. Bertino *A Conditional Purpose-based Access Control Model with Dynamic Roles*, Expert Systems with Applications, 38 (2011), pp. 1482–1489 (Q1, IF: 2.908, citations 18).
39. *Md. E. Kabir and **H. Wang**, Microdata Protection Method Through Microaggragation: A Median Based Approach, Information Security Journal: A Global Perspective Vol. 20(1), pp. 1-8, 2011 .
40. **H. Wang**, Y. Zhang, J. Cao. *Effective collaboration with information sharing in virtual universities.* IEEE Transactions on Knowledge and Data Engineering, 21(6), pp. 840-853, June, 2009 (Q1, IF: 2.07, citations 32)
41. L. Sun, **H. Wang**, J. Yong, *Authorization algorithms for permission-role assignments*, Journal of Universal Computer Science, Vol. 15, No. 9, pp. 1782-1796, 2009. (Q1, IF: 0.67)
42. **H. Wang**, J. Cao and Y. Zhang, Delegating revocations and authorizations in collaborative business environments, Information Systems Frontiers, Vol. 11(3): 293-305, 2009 (Q1, IF: 1.08)
43. *X. Sun, **H. Wang**, J. Li and T. Truta. *Enhanced P-sensitive K - anonymity models for privacy preserving data publishing*. Transactions on Data Privacy 1:2 (2008), pp: 53-66. (Q1, IF: 1.796, citations 39)
44. **H. Wang**, J. Cao, Y. Zhang, *Access control management for ubiquitous computing*, Future Generation Computer Systems journal (FGCS), Vol. 24, pp. 870-878, 2008 (Q1, IF: 2.79, citations 39)
45. *J. Li, **H. Wang**, J. Yong, *Current Developments of k-Anonymous Data Releasing*, Privacy and Security of Electronic Health Information, Vol. 3(1), pp. 109-120, 2008.
46. **H. Wang**, J. Cao, Y. Zhang, *A flexible payment scheme and its role based access control*, IEEE Transactions on Knowledge and Data Engineering (TKDE), 17(3): 425-436, 2005 (Q1, IF: 2.07, citations 74)
47. **H. Wang**, Y. Zhang, J. Cao, V. Varadharajan, *Achieving Secure and Flexible M-Services Through Tickets*, In B.Benatallah and Z. Maamar, Editor. IEEE Transactions on Systems, Man, and Cybernetics (TSMC). Part A, Vol. 33, Issue: 6, pages: 697- 708, Nov. 2003. **(Q1, IF: 6.22, citations 35)**
48. **H. Wang**, Y. Zhang, J. Cao, Y. Kambayahsi, *A global ticket-based access scheme for mobile users*, special issue on Object-Oriented Client/Server Internet Environments, Information Systems Frontiers, 6(1), pages: 35-46, 2004 (Q1, IF: 1.08, citations 15)
49. **H. Wang**, J. Cao, Y. Zhang, *An Electronic Payment Scheme and Its RBAC management*, Concurrent Engineering: Research and Application Journal, 12(3): 247- 257, Sept. 2004. (Q1, IF: 0.85, citations 9)
50. **H. Wang**. *A Signature Scheme with Controlled Number of Signatures*, International Journal Of Computer Engineering And Application, 1999, No.5. (Q1, IF: 3.72)
51. **H. Wang**. *The sufficient conditions of matrix*, Journal of Engineer Math. 1999, No.3 pp: 51-57. (Q1, IF: 0.8)

52. **H. Wang**. *The sufficient conditions of potentially stability linear system*, Journal of Application Mathematics, 1999, No. 2, pp: 31-37 (Q1, IF: 0.83)
53. **H. Wang**, Zh. You. *Potential stability for tridiagonal sign matrices*, Journal of Application Mathematics, 1997. No. 3, pp: 44-47. (Q1, IF: 0.83)
54. **H. Wang**, Zh. You. *Inverse H-matrices*, Quarterly Journal of Mathematics, 1996. 12, pp: 37-41. (Q1, IF: 0.64)
55. **H. Wang**, Zh. You. *Totally matrix*, Journal of Engineer Mathematics, 1995, No. 2, pp: 107-110 (Q1, IF: 0.8)
56. **H. Wang**, Ch. Wang. *On the Construction of Globally Asymptotically Stable Liyapunov Functions of a Type of Third- Order Systems with Nonseparable Variables*, Quarterly Journal of Mathematics, 1995, No. 2, pp: 94-105. (Q1, IF: 0.64)
57. **H. Wang**, Ch. Wang. *Conditions of LU as M-matrix*, Journal of Quarter Mathematics. 1994. No. 4, pp: 18-21. (Q1, IF: 0.64)

REFERRED CONFERENCE PAPERS:

58. *S. Subramani, **H. Wang**, etc. *Mining Actionable Knowledge Using Reordering Based Diversified Actionable Decision Trees*. WISE (1) 2016: 553-560.
59. *M. Peng, J. Huang, Z. Sun, S. Wang, **H. Wang**, G. Zhuo, G. Tian. *Improving Distant Supervision of Relation Extraction with Unsupervised Methods*. WISE (1) 2016: 561-568.
60. L. Sun, J. Ma, Y. Zhang, **H. Wang**. *Exploring Data Mining Techniques in Medical Data Streams.* ADC 2016: 321-332.
61. P. Vimalachandran, **H. Wang**, Y. Zhang, V. Wade and F. Whittaker. *Ensuring Data Integrity in Electronic Health Records: A Quality Health Care Implication.* Accepted by The 2016 IEEE International Conference on Orange Technologies (ICOT 2016), Australia.
62. T. Ma, **H. Wang**, J. Cao, J. Yong, Y. Zhao. *Access control management with provenance in healthcare environments*. CSCWD 2016: 545-550
63. Supriya, Siuly, **H. Wang**, G. Zhuo, Y. Zhang. *Analyzing EEG Signal Data for Detection of Epileptic Seizure: Introducing Weight on Visibility Graph with Complex Network Feature*. ADC 2016: 56-66.
64. *W. Yao, J. He, **H. Wang** and Y. Zhang, *Collaborative Topic Ranking: Leveraging Item Meta-data for Sparsity Reduction*, the 29th AAAI Conference on Artificial Intelligence (AAAI-15) (2015).
65. *M. Peng, J. Zhu, X. Li, J. Huang, **H. Wang**, and Y. Zhang. *Central Topic Model for Event-oriented Topics Mining in Microblog Stream.* CIKM '15, 1611-1620
66. *J. Huang, M. Peng, and **H. Wang**. Topic Detection from Large Scale of Microblog Stream with High Utility Pattern Clustering. PIKM '15. 3-10. DOI=http://dx.doi.org/10.1145/2809890.2809894
67. Y. Qin, **H. Wang**, J. Zhang, X. Tao, W. Zhang, K. Taylorx and Q. Sheng, Efficient Algorithms for Scheduling XML Data in a Mobile Wireless Broadcast Environment, Accepted by ICPADS 2015.
68. T. Ma, **H. Wang**, J. Yong and Y. Zhao. Causal dependencies of provenance data in healthcare environment. *Computer Supported Cooperative Work in Design (CSCWD), 2015 IEEE 19th International Conference on*, Calabria, 2015, pp. 643-648.
69. Y. Qin, Q. Sheng, N. Falkner, W. Zhang, **H. Wang**. *Indexing Linked Data in a Wireless Broadcast System with 3D Hilbert Space-Filling Curves*. CIKM 2014: 1775-1778 .
70. G. Wang, **H. Wang**, X. Tao, J. Zhang, X. Yi, J. Yong. *Positive Influence Dominating Set Games*. CSCWD 2014: 469-473.
71. Y. Qin, Q. Sheng, N. Falkner, S. Dustdar, **H. Wang**, A. Vasilakos. *When Things Matter: A Data-Centric View of the Internet of Things*. CoRR abs/1407.2704 (2014)
72. J. Zhang, **H. Wang**, X. Tao. *SODIT: An Innovative System for Outlier Detection using Multiple Localized Thresholding and Interactive Feedback*, ICDE2013, pp. 1364-1367.
73. L. Sun, **H. Wang** and E. Bertino. *Role-based access control to outsourced data in cloud computing*. Proceedings of ADC2013, pp. 80-89, Adelaide, Australia.
74. Y. Qin, Q. Sheng, M. Mehdi, **H. Wang**, D. Xie. *Effectively Delivering XML Information in Periodic Broadcast Environments.* DEXA (1) 2013: 165-179
75. **H. Wang**, J. Zhang, J. Soar, X.Tao, W. Huang. *SIMS: Self-adaptive Intelligent Monitoring System for Supporting Home-Based Heart Failure Patients*. ICOST 2013: 322-330.
76. G. Wang, **H. Wang**, X. Tao, J. Zhang, J. Zhang. *Minimising K-Dominating Set in Arbitrary Network Graphs*. ADMA (2) 2013: 120-132.
77. G. Wang, **H. Wang**, X. Tao, J. Zhang. *A Self-Stabilizing Algorithm for Finding a Minimal Positive Influence Dominating Set in Social Networks*. ADC 2013: 93-99.
78. *M. Baig, J. Li, J. Liu, X. Ding, **H. Wang**: Data Privacy against Composition Attack. DASFAA (1) 2012: 320-334
79. G. Wang, **H. Wang**, X. Tao, J. Zhang. A Self-stabilizing Algorithm for Finding a Minimal K-Dominating Set in General Networks. ICDKE 2012: pp. 74-85.
80. X. Tao, Y. Li, R. Lau, **H. Wang**. *Unsupervised Multi-label Text Classification Using a World Knowledge Ontology*. PAKDD (1) 2012, pp. 480-492

81. M. Baig, J. Li, J. Liu, X. Ding, **H. Wang**. *Data Privacy against Composition Attack*. DASFAA (1) 2012, pp. 320-334
82. L. Sun, **H Wang**, X. Tao, Y. Zhang, and J. Yang. *Privacy Preserving Access Control Policy and Algorithms for Conflicting Problems*. In Proceedings of the 10th International Conference on Trust, Security and Privacy in Computing and Communications (TRUSTCOM '11), pp. 250-257.
83. L. Sun, **H. Wang**. *A purpose based usage access control model for E-healthcare services*. ICDKE 2011, pp. 41-46.
84. G. Wang, **H. Wang**, X. Tao, J. Zhang. *Positive Influence Dominating Set in E-Learning Social Networks*. ICWL 2011, pp. 82-91
85. Y. Qin, **H. Wang**, L. Sun. *Cluster-Based Scheduling Algorithm for Periodic XML Data Broadcast in Wireless Environments*, AINA Workshops, pp. 855-860, 2011
86. M. Baig, J. Li, J. Liu, X. Ding, **H. Wang**. *Cloning for privacy protection in multiple independent data publications*. CIKM 2011, pp. 885-894
87. *M.E. Kabir, **H. Wang** and Y. Zhang. *A Pairwise-Systematic Microaggregation for Statistical Disclosure Control*, ICDM 2010, pp. 266-273. 2010.
88. J. Zhang, S. Dekeyser, **H. Wang** and Y. Shu. *DISTRO: A System for Detecting Global Outliers from Distributed Data Streams with Privacy Protection*. DASFFA 2010, pp. 477-481, 2010.
89. Y. Qin, **H. Wang**. *Effectively Scheduling XML Data for On-Demand Broadcasts in Wireless Environment*, ADC 2011, pp. 95-102, 2011.
90. J. Zhang, **H. Wang** and S. Dekeyser, *On Memory and I/O Efficient Duplication Detection for Multiple Self-clean Data Sources,* Proceedings of the 3rd International Workshop on Managing Data Quality in Collaborative Information Systems (with DASFAA'01 workshop), pp. 130-142, 2010.
91. X. Sun, **H. Wang** and J. Li **,** *Satisfying Privacy Requirements: One Step Before Anonymization*, PAKDD2010, pp. 181-188, 2010.
92. *M. Kabir, **H. Wang**, E. Bertino and Y. Chi, *Systematic Clustering Method for l-diversity Model*, Brisbane, Queensland, Australia. CRPIT (ADC 2010), 104. Shen, H. T. and Bouguettaya, A., Eds. ACS. pp. 91-101.
93. M. Li, X. Sun, **H. Wang** and Y. Zhang, *Optimal Privacy-aware Path in Hippocratic Databases*, *DASFAA2009*, pp. 441-455, 2009.
94. X. Sun and **H. Wang** and J. Li, *Injecting Purpose and Trust into Data Anonymisation*, CIKM 2009, pp. 1541-1544. 2009.
95. **H. Wang**, J. Cao, Y. Zhang, *Formal Authorization Allocation Approaches for Role-Based Access Control Based on Relational Algebra Operations*, WISE'2002, pp. 301-310, 2002.
96. **H. Wang**, J. Cao, Y. Zhang, *A Consumer Anonymity Scalable Payment Scheme with Role Based Access Control*, WISE'2001, pp. 53-62, 2001.

ARC grants awarded in the last 10 years

Project ID	CI Names	Amount	Project title	Period and Publications
DP0663414	H. Wang	$97,000	Protect information sharing within distributed collaborative environment	2006-2008, [41, 42, 43, 44, 93-94]
DP0774450	J. Li and H. Wang	$165,000	Privacy preserving data sharing in data mining environments	2007-2009, [45, 46, 48, 49, 50]
DP0988465	H. Wang	$245,000	Limiting disclosure of private information in relational database systems	2009-1011, [15, 17, 18, 20, 21-25, 27-31, 81, 83, 84, 86]
LP150100673	Zhang, Wang, Yi, Soar, Miao, Ben	$295,000	Privacy Preserving Data Sharing in Electronic Health Environment	2015-2017, [1, 3, 4, 67, 68]

F14. Research Opportunity and Performance Evidence (ROPE) - Ten career-best research outputs

(Please upload a PDF with a list of your ten career-best research outputs, with a brief paragraph for each research output explaining its significance (five pages maximum).)

Uploaded PDF file follows on next page.

Ten career-best publications – Professor Hua Wang

All 10 papers are in Q1 ranking (http://www.scimagoir.com/, Journal Rankings) and with impact factors information.

1. *J. Ma, L. Sun, **H. Wang**, Y. Zhang and U. Aickelin, Supervised Anomaly Detection in Uncertain Sensor Data Streams. *ACM Transactions on Internet Technology*. 16(1): 67-86, 2016. DOI = http://dx.doi.org/10.1145/2806890

Impact Factor: 1.72 (Q1). This paper presents a novel framework that supports anomaly detection in uncertain data streams. The proposed framework adopts the wavelet soft-thresholding method to remove the noises or errors in data streams.

2. *J. Huang, M. Peng, **H. Wang**, J. Cao, G. Wang, X. Zhang. A Probabilistic Method for Emerging Topic Tracking in Microblog Stream. *World Wide Web*, 20(2): 325-350, 2017. (Q1, IF: 1.47)

Impact Factor: 1.76 (Q1). This paper proposes a novel emerging topics tracking method, which aligns emerging word detection from temporal perspective with coherent topic mining from spatial perspective.

3. Supriya, Siuly, **H. Wang**, J. Cao and Y. Zhang. Weighted visibility graph with complex network features in the detection of epilepsy. *IEEE Access*, Vol. 4, pp: 6554 – 6566, 2016.

Impact Factor: 1.27 (Q1). This paper was the collaborative work of the three CIs that introduces a new idea for epilepsy detection using complex network statistical properties by measuring different strengths of the edges in natural visibility graph theory, which is considered as weight. The epileptic EEG signals are transformed into complex network and then two important statistical properties of a network such as modularity and average weighted degree used for extracting the imperative characteristics from a network of EEG signals.

4. ***H. Wang**, L. Sun and E. Bertino. Building access control policy model for privacy preserving and testing policy conflicting problems. *Journal of Computer and System Sciences*. 80 (8): 1493–1503, 2014.

Impact Factor: 1.12 (Q1). This paper analyses the privacy preserving models and evaluated the access methods for outsourcing dataset.

5. *X. Sun, **H. Wang** and J. Li. Publishing anonymous survey rating data. *Data Mining and Knowledge Discovery*. 23(3): 379-406, 2011. Outcomes from ARC DP0774450 (13 citations).

Impact Factor: 1.99 (Q1). This paper remedied the problem by defining a novel principle called (k, ϵ)-anonymity model to protect privacy. The anonymity model is formulated by its graphical representation and a specific graph-anonymisation problem is studied by adopting graph theory.

6. Y. Zhang, Y. Shen, **H. Wang**, Y. Zhang, X. Jiang. On Secure Wireless Communications for Service Oriented Computing. *IEEE Transactions on Services Computing*, No. 1, pp. 1, 2015. Preprints, DOI = 10.1109/TSC.2015.2478453

Impact Factor: 3.05 (Q1). This paper establishes a theoretical framework for the study of eavesdropper-tolerance capability in a two-hop wireless network, where the cooperative jamming is

adopted to ensure security defined by secrecy outage probability and opportunistic relaying is adopted to guarantee reliability defined by transmission outage probability.

7. **H. Wang**, J. Cao, Y. Zhang. A flexible payment scheme and its role based access control. *IEEE Transactions on Knowledge and Data Engineering (TKDE).* 17(3): 425-436, 2005 (74 citations).

Impact Factor: 2.07 (Q1). This paper with the three CIs proposes a practical payment protocol with scalable anonymity for Internet purchases, and analyzes its role-based access control management.

8. **H. Wang**, Y. Zhang, J. Cao. Effective collaboration with information sharing in virtual universities. *IEEE Transactions on Knowledge and Data Engineering*. 21(6): 840-853, 2009. Outcomes from ARC DP0663414 (40 citations).

Impact Factor: 2.07 (Q1). This paper was a collaboration of the three CIs that presents a new rule-based framework to identify challenges of sharing in virtual university environments through role-based access control.

9. **H. Wang**, J. Cao, Y. Zhang. Access control management for ubiquitous computing. *Future Generation Computer Systems*. Vol. 24: 870-878, 2008. Outcomes from ARC DP0663414 (45 citations)

Impact Factor: 2.79 (Q1). This paper with the three CIs discusses a usage control model to protect services and devices in ubiquitous computing environments, which allows the access restrictions directly on services and object documents.

10. Y. Zhang, Y. Shen, **H. Wang**, and X. Jiang, On Secure Wireless Communications for IoT under Eavesdropper Collusion, IEEE Transactions on Automation Science and Engineering, vol.13, no.3, 1281-1293, 2016

Impact Factor: 2.696 (Q1). The paper studies the important secrecy outage performance of wireless communications under eavesdropper collusion, where the physical layer security is adopted to counteract such attack. Based on the classical Probability Theory, the authors first conduct analysis on the secrecy outage of the simple noncolluding case in which eavesdroppers do not collude and operate independently.

F15. Research Opportunity and Performance Evidence (ROPE) - Further evidence in relation to research impact and contributions to the field, including those most relevant to this Proposal

(Write a maximum of 11250 characters (approximately 1500 words). Detail further evidence in relation to research impact and contributions to the field. Click on the information icon and refer to the Instructions to Applicants for the required content and formatting.)

(1) Hua has won three ARC Discovery Projects and one Linkage Project since 2006:
• LP150100673, Privacy Preserving Data Sharing in Electronic Health Environment, (Y.Zhang, H. Wang et al, $295,000 + $150,000, 2016-2018)
• DP0988465. H. Wang. Limiting disclosure of private information in relational database systems, $245000, 2009-2011.
• DP0774450. J. Li and H. Wang. Privacy preserving data sharing in data mining environments, $186000, 2007-2009.
• DP0663414. H. Wang. Protect information sharing within distributed collaborative environment, $98000, 2006-2008.

With these four ARC grants, many peer reviewed research papers have been published by journals and international conferences. 7 PhD students, supported by the grants, have already graduated. Furthermore many papers are used as study materials for postgraduate students.

(2) Hua has served as an ARC Reader from 2009.

(3) Hua served as the Co-Editor-in-Chief, ICST Transaction on Scalable Information Systems (http://icst.org/scalable-information-systems/), 2009 onwards.

(4) He was the Guest editor for a special issue on “Security and Privacy in Cloud Computing” for The Computer Journal (Ranked “A*” by the ERA-2010).

(5) He was the Guest editor of a special issue on “Data and Knowledge Engineering in open social network” for Journal of Software (Ranked “B” by the ERA), 2010.

(6) Hua was the Guest editor of a special issue on “RFID security and privacy” for International Journal of Security and Networks (IJSN), 2009.

(7) Hua was an associate member of the ARC Research Network in Enterprise Information Infrastructure (EII) and also a member of ARC networks on security.

(8) Hua has served as an editorial board member of "The Open Cybernetics and Systemics Journal". 2007 onwards

In addition, Hua has been served as chairs for 50+ international workshops/conferences in the past 10 years such as:

• The 5th International Conference on Health Information Science (HIS 2016), Shanghai, China, November 6-7, 2016 (Conference Organising Chair)
• The 4th International Conference on Health Information Science (HIS 2015), Melbourne, Australia, May, 2015 (Conference Organising Chair)
• The 17th International Conference on Web Information Systems Engineering (WISE 2016), Shanghai, China, November 8-10, 2016 (Publication Chair)
• The 16th International Conference on Web Information Systems Engineering (WISE 2015),Florida, USA, November 1-3, 2015 (Publication Chair)
• The 25th Australasian Database Conference (ADC 2014), Brisbane, Australia, July, 2014 (General Chair)
• The 24th Australasian Database Conference (ADC 2013), Adelaide, South Australia, 29 January - 1 February, 2013 (General Chair)
• The 29th International Conference on Data Engineering (ICDE 2013), workshop, Australia (PC member)
• The International Conference on Data and Knowledge Engineering (ICDKE), Fujian, China, November 21-23, 2012 (General Chair)
• The 14th Asia-Pacific Web Conference (APWeb), April 11-13, 2012, Kunming, China (Workshop Co-Chair)
• The 35 Australasian Computer Science Conference (ACSC 2012), Melbourne, Australia. January 30 - February 2, 2012. (PC member)
• The 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-11), Changsha, China, November 16-18, 2011 (Program Chair)

• The 10th International Conference on Web-based Learning (ICWL2011), December 8-10, 2011, Hong Kong (Publicity Co-Chair)
• The 11th International Conference on Web Information Systems Engineering (WISE 2010), December 12-14, 2010, Hong Kong, China (Publicity Co-Chairs)

(2) Hua has published 166 refereed publications with over 1400 citations in the last 10 years (https://scholar.google.com.au/citations). His published papers are started to be used for postgraduate students as study materials. For example, the following paper is used for Masters' students at the University of Southern Queensland to show how to use Statistical methods for privacy preserving :

Kabir, M.E., Wang, H., and Yanchun, Z., A Pairwise-Systematic Microaggregation for Statistical Disclosure Control, Proceedings of 10th International Conference on Data Mining (ICDM 2010), Sydney, Australia.

F16. Currently held ARC Projects

(This information is auto-populated from your RMS profile and will include any 'active' Project which has not yet had a Final Report approved and the Project file closed by the ARC. If you have any concerns with the information recorded here, contact your Administering Organisation's Research Office. NOTE: If you hold a CI or a PI role on the Project/s listed in the table below you must ensure a progress statement is provided in G2. This requirement applies to the following schemes: Discovery Projects, Discovery Indigenous Researchers Development, Discovery Indigenous, Discovery Early Career Researcher Award, Linkage Projects, Industrial Transformation Research Hubs, Industrial Transformation Training Centres or any ARC Fellowship scheme. Please click on the information icon and refer to the Instructions to Applicants for further information.)

Identifier	Scheme Name	Investigators	Admin Organisation	Project Title	Funding	End Date
LP150100673	LP 2015 R1	Prof Yanchun Zhang ; Prof Hua Wang ; Prof Xun Yi ; Prof Jeffrey Soar ; Prof Yuan Miao ; Mr Ben Heyward	Victoria University	Privacy Preserving Data Sharing in Electronic Health Environment	$295,467	14/06/2019

Part F - Personnel and ROPE (Prof Yanchun Zhang)

F1. Personal Details

(To update personal details, click the 'Manage Personal Details' link below. Note this will open a new browser tab. When returning to the form ensure you 'Refresh' the page to capture the changes made to your profile.)

Participation Type

Chief Investigator

Title

Prof

First Name

Yanchun

Second Name

Family Name

Zhang

F2. Will you be residing predominately in Australia for the duration of the Project?

(This is a 'Yes' or 'No' question. Indicate whether you will be residing predominantly in Australia for the duration of the Project, taking into account any international travel. If you are applying as a CI and you answer 'No' to this question you will be prompted to contact your Research Office to check your eligibility.)

Yes

F3. Qualifications

(To update any qualifications, click on the 'Manage Qualifications' link below. Note this will open a new browser tab. When returning to the form ensure you 'Refresh' the page to capture the changes made to your profile.)

Conferral Date	AQF Level	Degree/Award Title	Discipline/Field	Awarding Organisation	Country of Award
08/11/1991	Doctoral Degree	PhD	Computer Science	The University of Queensland	Australia
20/02/1982	Bachelor Degree	BSc	Computer Science	Hebei University	China (excludes SARs and Taiwan)

F4. Are you currently undertaking a Higher Degree by Research which will be conferred after 1 January 2018?

(This is a 'Yes' or 'No' question. If you are applying as a CI and your answer is 'Yes' to this question you will be prompted to contact your Research Office.)

No

F5. Research Opportunity and Performance Evidence (ROPE) – Current and previous appointment(s) / position(s) - during the past 10 years

(To update any details in this table, click on the 'Manage Employment Details' link below. Note this will open a new browser tab. When returning to the form ensure you 'Refresh' the page to capture the changes made to your profile.

Click on the information icon above and refer to the Instructions to Applicants for more information.)

Description	Department	Contract Type	Employment Type	Start Date	End Date	Organisation
Professor of Computer Science	College of Engineering and Science	Permanent	Full Time	01/01/2005		Victoria University

F6. Employment Details as at Commencement date of Project

(This question will be used to determine your eligibility. Confirm your employment status at all organisations that you will be associated with as at the Commencement Date for the Project (1 January 2018). Enter the relevant appointment type and Full-Time Equivalent (FTE) for each organisation. Click on the information icon for further information.)

Org name	Is this an Eligible Organisation?	Please choose your appointment type for this organisation.	Please enter your FTE for this Organisation
Victoria University	Yes	Employee	1

F7. Further Details Regarding Partner Investigator Status - Do you hold a remunerated appointment at an Eligible Organisation?

(NOTE: this question is mandatory ONLY FOR PIs WHO:
• at F2 confirmed that they will reside predominantly in Australia for the duration of the proposed Project; AND
• at F4 confirmed that they are not currently undertaking a Higher Degree by Research which will be conferred after 1 January 2018; AND
• at F6 indicated that they would hold either:
- an appointment at an Eligible Organisation equal or greater than 0.2FTE; OR
- Emeritus appointment at an Eligible Organisation

This is a 'Yes' or 'No' question. If you select 'Yes', you will be further prompted to justify your participation on this Proposal as a PI with reference to sections F7.2 and F7.3 of the Funding Rules. Click on the information icon for further information.)
Do you hold a remunerated appointment at an Eligible Organisation?

Justification of PI status

F8. Relevant Organisation for this Proposal

(Enter the Organisation that is relevant to your participation on this proposal, and that you will be associated with as at the Commencement Date for the Project (1 January 2018). The 'relevant organisation' is the primary organisation that will be supporting your involvement in this Project if it is funded. Note that the Organisation must be listed in F6 for this question to validate.)

Relevant Organisation

Victoria University

F9. What is your time commitment to this Project?

(Enter your time commitment to this Project as a Full-Time Equivalent (FTE). Note that a FTE of 1.0 represents a full-time commitment (i.e. 5 days per week).)

0.2

F10. Are you applying for Teaching Relief?

(This is a 'Yes' or 'No' question.

• If you answer 'Yes' to this question a budget line will be automatically populated for the Teaching Relief in the budget table in Form Part D: Project Cost. This will allow you to enter the funding amount requested in the relevant year/s. To remove the Teaching Relief from the budget table you must return to this question and answer 'No'.

• Note: CIs may request funding for teaching relief or other duties in order to maximise the opportunity for the CI to conduct research. This question is only relevant for CIs and will not be activated for PIs. If the answer to this question is 'yes', the budget table in 'Project Costs' will automatically update with a line to add costs for Teaching Relief.)

No

F11. Are you applying for a Discovery International Award?

(This is a 'Yes' or 'No' question. If you answer 'Yes' to this question a budget line will be automatically populated for the DIA in the budget table in Form Part D: Project Cost. This will allow you to enter the funding amount requested in the relevant year/s. To remove the DIA from the budget table you must return to this question and answer 'No'. Click on the information icon for further information.)

No

F12. Research Opportunity and Performance Evidence (ROPE) - Details of your career and opportunities for research

(Write a maximum of 5250 characters (approximately 750 words). Please detail your career and opportunities.)

(i) Prof. Zhang has over 20 years' extensive research experience, since he received his PhD degree of Computer Science from The University of Queensland in 1991.

(ii) Prof. Zhang is a full professor and a founding director of Centre for Applied Informatics in Victoria University. He is a world-leading expert in data mining, information system, Web & Internet technology, sensor networks and applied informatics. His outstanding research performance is evidenced by over 280 publications in international journals and refereed conferences proceedings. He has authored/edited a dozen books/proceedings and journal special issues in the related areas. His research has been widely cited (over 10,000 citations according to Google Scholars as in 2016, H-index: 36). He was a member of Australian Research Council (ARC) College of Experts (2008-2010).

Prof. Zhang has been leading a team in developing e-Research program. He has made strong impacts through his leadership in applied research and collaborations with universities and industry organizations such as UQ, SEQ Healthy Waterways Partnership (HWP) and Royal Brisbane and Women's Hospital. He has 9 ARC grants in data mining, database and multidisciplinary e-research fields since 2003. For example, his current three projects: Using Data Mining Methods to Remove Uncertainties in Sensor Data Streams, Real-time and Self- Adaptive Stream Data Analyser for Intensive Care Management, and Data Exchange and Service Integration with Applications in Health Information Systems attracted $930K of ARC funding. Also, recent collaborations with UQ and HWP in ARC Linkage Project: Data Enhancement, Integration and Access Services for Smarter, Collaborative and Adaptive Whole-of-Water-Cycle Management have achieved high quality outcomes published on Neurocomputing Journal (2012, ERA B, Impact Factor: 1.634).

Upon submitting his PhD thesis, he was employed at UQ and CRC: Distributed Systems Technology Centre as a research fellow from Aug. 1991 to Feb. 1994, undertaking research in the areas of distributed databases and collaborative information systems. From 1994 to June 2003, he worked as a lecturer, senior lecturer and associate professor at the University of Southern Queensland. In June 2003, he joined Victoria University as an associate professor, soon promoted to full professor in 2004. At USQ/VU, he performed the varied duties: teaching, research, and administration with various time distributions. Since 2010, he has taught one unit per year, the remaining time is dedicated to carrying research, leading the research centre, and supervising PhD students (currently 5 full time PhD students). More time is allocated to research and research leadership, with the time distribution of research, teaching and administration around (60%, 20%, 20%) in the last 5 years. He has won

VU Vice Chancellor's Peak Award for Research and Research Training (Research Supervision) in 2011.

(iii) Prof. Zhang is a full time academic (research and teaching).

(iv) He has no career interruption.

(v) Victoria University has a world class research environment in fields of Artificial Intelligence and Image Processing and Distributed Computing, according to Excellence in Research for Australia (ERA) 2012 evaluation; these are two major research fields of CAI. VU has suitable mentoring facilities for applied informatics research, due to the University support of Applied Informatics as one of VU's strategic research areas or strength. (vi) He is Founding Editor and Editor-In-Chief of both World Wide Web Journal (WWWJ) (ERA A journal) from Springer and Health Information Science and System Journal (HISS) from BioMed Central. He is the initiator of International Conference of Web Information Systems Engineering (WISE) (ERA A conference), and he has been the steering committee chair since 2000. He is also the Editor of Web Information Systems Engineering and Internet Technology Book Series from Springer, the Chairman of International Web Information Systems Engineering Society (WISE Society) and Australian representative of International Federation of Information Processing (IFIP) 's Working Group WG6.4 on Internet Applications Engineering. His research especially focuses on applying privacy preserving approaches and data mining to aid efficient detect events and patterns from environments and assist health care. The team has lots of collaborative research publications, representative work includes

(1) Supriya, Siuly, H. Wang, J. Cao and Y. Zhang. Weighted visibility graph with complex network features in the detection of epilepsy. IEEE Access, Vol. 4, pp: 6554 - 6566, 2016. (Q1; IF=1.270)

(2) H. Wang, J. Cao, Y. Zhang. A flexible payment scheme and its role based access control. IEEE Transactions on Knowledge and Data Engineering (TKDE). 17(3): 425-436, 2005 (Q1; IF-2.07, 74 citations).

(3) H. Wang, Y. Zhang, J. Cao. Effective collaboration with information sharing in virtual universities. IEEE Transactions on Knowledge and Data Engineering. 21(6): 840-853, 2009. Outcomes from ARC DP0663414 (Q1; IF-2.07, 40 citations).

F13. Research Opportunity and Performance Evidence (ROPE) - Significant research outputs and ARC grants

(Upload a PDF of no more than four A4 pages with a list of all research outputs, such as journal articles and refereed conference papers, book and book chapters. Use asterisks to identify research outputs relevant to this Proposal. Click on the information icon or refer to the Instructions to Applicants for the required content and formatting.)

Uploaded PDF file follows on next page.

Recent significant publications and ARC grants: Yanchun Zhang

Prof. Yanchun Zhang has built a sound and solid record in Data Mining, E-Research (E-Health), Social Networking and Web service fields. He has over **280** scholar papers in these fields published by high quality journals and top international conferences such as: ACM Transactions on Internet Technology, Information Systems, ACM Transactions On Computer-Human Interaction, SIGMOD, AAAI and ICDE. His publications are widely cited (over **10,000** citations according to Google Scholars in 2016) and his H-index is **36**.
(Journal Rankings are based on SciMagoir at http://www.scimagoir.com/)

REFEREED JOURNAL PAPERS:

1. Y. Zhang, Y. Shen, H. Wang, **Y. Zhang**, X. Jiang. *On Secure Wireless Communications for Service Oriented Computing*. IEEE Transactions on Services Computing, doi:10.1109/TSC.2015.2478453 (Q1, IF: 3.05**)**
2. *Supriya, Siuly, H. Wang, J. Cao and **Y. Zhang**. *Weighted visibility graph with complex network features in the detection of epilepsy*. IEEE Access, Vol. 4, pp: 6554 – 6566, 2016. (Q1, IF: 1.270)
3. J. Ma, L. Sun, H. Wang, **Y. Zhang** and U. Aickelin, *Supervised Anomaly Detection in Uncertain Sensor Data Streams.* ACM Transactions on Internet Technology (TOIT), 16, 1, Article 4 (January 2016), 20 pages. DOI= http://dx.doi.org/10.1145/2806890 (Q1, IF: 1.72)
4. *Siuly, H. Wang and **Y. Zhang**. *Detection of motor imagery EEG signals employing Naïve Bayes based learning process*. Measurement. Volume 86: 148–158, 2016 (Q2, IF: 1.48)
5. D. Pandey, X. Yin, H. Wang, **Y. Zhang**. *Accurate Vessel Segmentation using Maximum Entropy incorporating Line Detection and Phase-Preserving Denoising*, Accepted by Computer Vision and Image Understanding on December, 2016. (Q1, IF: 2.134)
6. *S. Siuly, E. Kabir, H. Wang, and **Y. Zhang**. *Exploring Sampling in the Detection of Multicategory EEG Signals*, Computational and Mathematical Methods in Medicine, vol. 2015, Article ID 576437, 12 pages, 2015. doi:10.1155/2015/576437 (Q1, IF: 0.98)
7. L. Sun, J. Ma, H. Wang, and **Y. Zhang**. *Cloud Service Description Model: An Extension of USDL for Cloud Services*, IEEE Transactions on Services Computing. 2015, doi: 10.1109/TSC.2015.2474386 (Q1, IF: 3.05**)**
8. *G. Huang, J. He, **Y. Zhang**, W. Zhou, H. Liu, P. Zhang, Z. Ding, Y. You, J. Cao. *Mining Streams of Short Text for Analysis of World-Wide Event Evolutions*, World Wide Web, Vol. 18, No. 1201—1217, 2015. (Q1, IF: 1.47)
9. *X. Yin, S Hadjiloucas, **Y Zhang,** J. He, Y. Wang and D. Zhang. *Application of Complex Extreme Learning Machine to Multiclass Classification problems with High Dimensionality: a THz spectra classification problem*, Digital Signal Processing, 2015, DOI: 10.1016/j.dsp.2015.01.007. (Q1, IF: 1.44)
10. X. Zhou, J. He, G. Huang, **Y. Zhang**. *SVD-based incremental approaches for recommender systems*. *J. Comput. Syst. Sci.* 81(4): 717-733 (2015). (Q1, IF: 1.58)
11. X. Yin, S Hadjiloucas, and **Y Zhang**. *Complex extreme learning machine applications in terahertz pulsed signals feature sets.* Computer Methods and Programs in Biomedicine, 117 (2), pp 387-403, 2014. (Q1, IF: 1.86).
12. X. Yin, B. W.-H. Ng, J. He, **Y. Zhang**, and D. Abbott. *Accurate image analysis of the retina using Hessian matrix and binarisation of thresholded entropy with application of texture mapping*. PLoS One, 9(4), e95943, 2014. (Q1, IF: 3.53).
13. Z. Xie, G. Huang, J. He, **Y. Zhang**. *A Clique-based WBAN Scheduling for Mobile Wireless Body Area Networks*. Journal of Procedia Computer Science, Volume 31, Pages 1092–1101 (2014).
14. Z. Xie, G. Huang, R. Zarei, J. He, **Y. Zhang**. *Wireless Sensor Networks for Heritage Deformation Detection and Tracking*, Sensors, 14, 20562-20588 (2014).
15. X. Zhou, J. He, G. Huang, **Y. Zhang**. *Scalable approximating SVD algorithm for recommender systems*. Web Intelligence and Agent Systems 12(4): 359-373 (2014).
16. *G. Huang, **Y. Zhang**, J. Cao, M. Steyn, K. Taraporewalla. *Online mining abnormal period patterns from multiple medical sensor data streams.* World Wide Web 17(4): 569-587 (2014). (Q1, IF: 1.47)
17. Q. Ma, Y. Luo, P. Guo, G. Gao, M. Yang, G. Sablok, **Y. Zhang**, F. Zhou. *Clinical Effects of Xinmailong Therapy in Patients with Chronic Heart Failure.* Int. J. Med. Sci. 2013; 10(5): 624-633. doi:10.7150/ijms.5779
18. H. Zhang, J. Liu, **Y. Zhang**. *Utilizing BDI Agents and a Topological Theory for Mining Online Social Networks*. Fundam. Inform. 127(1-4): 479-494 (2013).
19. Z. Wu, G. Xu, C. Lu, E. Chen, **Y. Zhang**, H. Zhang. *Position-wise contextual advertising: Placing relevant ads at appropriate positions of a web page*. Neurocomputing 120: 524-535 (2013).
20. X. Yi, **Y. Zhang**. *Equally contributory privacy-preserving k-means clustering over vertically partitioned data*. *Inf. Syst.* 38(1): 97-107 (2013). (Q1, IF: 2.07).
21. Y. Yang, Y. Yang, H. Shen, **Y. Zhang**, X. Du, X. Zhou. *Discriminative Nonnegative Spectral Clustering with Out-of-Sample Extension*. IEEE Trans. Knowl. Data Eng. 25(8): 1760-1771 (2013). (Q1, IF: 2.35)

22. L. Li, G. Xu, Z. Yang, P. Dolog, **Y. Zhang**, M. Kitsuregawa. *An efficient approach to suggesting topically related web queries using hidden topic model.* World Wide Web 16(3): 273-297 (2013). (Q1, IF: 1.47)
23. J. Cao, Z. Wu, B. Mao, **Y. Zhang**. *Shilling attack detection utilizing semi-supervised learning method for collaborative recommender system.* World Wide Web 16(5-6): 729-748 (2013). (Q1, IF: 1.47)
24. X. Tao, Y. Miao and **Y. Zhang.** *Cooperative-Competitive Health care Service Negotiation*, International Journal of Software and Informatics, 6(4), 2012.
25. *J. He, **Y. Zhang**, and G. Huang. *CIRCE: Correcting Imprecise Readings and Compressing Excrescent Points for Querying Common Patterns in Uncertain Sensor Streams*, Information Systems 38(8): 1234-1251, 2013 available online 25 Jan., 2012. (Q1, IF: 1.595).
26. J. He, **Y. Zhang,** G. Huang, Y. Shi and J. Cao. *Distributed Data Possession Checking for Securing Multiple Replicas in Geographically Dispersed Clouds*, Journal of Computer and System Sciences, 78(5): 1345-1358, 2012. (Q1, IF: 1.631)
27. J. He, **Y. Zhang,** G. Huang and J. Cao. *A Smart Web Service based on the Context of Things*, ACM Transactions on Internet Technology, 11(3), 2012. (Q1, IF: 1.118)
28. J. He, **Y. Zhang**, and G. Huang. *Exceptional Object Analysis for Finding Rare Environmental Events from Water Quality Datasets*, Neurocomputing Journal, 92(1): 69-77, Sep. 2012. (Q1, IF:1.429)
29. G. Huang, **Y. Zhang**, J. He and J. Cao. *Fault Tolerance in Data Gathering Wireless Sensor Networks*, The Computer Journal 54(6): 976-987, 2011. (Q1. IF: 1.327)
30. Y. Hao, **Y. Zhang** and J. Cao. *A Novel QoS Model and Computation Framework in Web Service Selection*. World Wide Web. Accepted on Jan. 5, 2012 (Q1, IF: 1.47)
31. L. Li, G. Xu, Z. Yang, **Y. Zhang** and M. Kitsuregawa. *An Efficient Approach to Suggesting Topically Related Web Queries Using Hidden Topic Model*, World Wide Web. (Q1, IF: 1.47)
32. Z. Wu, G. Xu, P. Dolog**, Y. Zhang**. *An Improved Contextual Advertising Matching Approach based on Wikipedia Knowledge,* The Computer Journal, June 2011.
33. X. Sun, H. Wang, J. Li and **Y. Zhang**. *Satisfying Privacy Requirements Before Data Anonymization*, The Computer Journal, 30 (5): 332-345, 2011.
34. X. Sun, H. Wang, **Y. Zhang**. *On the identity anonymization of high-dimensional rating data*. Concurrency and Computation: Practice and Experience 24(10): 1108-1122 (2012)
35. X. Sun, H. Wang, J. Li, **Y. Zhang**. *An approximate microaggregation approach for microdata protection.* Expert Syst. Appl. 39(2): 2211-2219 (2012)
36. G. Huang, M. Hu, Y. Zhou, P. Liu, **Y. Zhang**. *A Distributed Trust Model Based on Reputation Management of Peers for P2P VoD Services.* TIIS 6(9): 2285-2301 (2012)
37. M. Li, X. Sun, H. Wang, **Y. Zhang**, J. Zhang. *Privacy-aware access control with trust management in web service*, World Wide Web, Vol. 14, No. 3. (2011). (Q1, IF: 1.47).
38. M. Li, X. Sun, H. Wang, **Y. Zhang.** *Multi-level delegations with trust management in access control systems*. J. Intell. Inf. Syst. 39(3): 611-626 (2012)
39. Z. Wu, G. Xu, **Y. Zhang**, Z. Cao, G. Li, Z. Hu. *GMQL: A graphical multimedia query language.* Knowl.-Based Syst. 26: 135-143 (2012)
40. Z. Wu, G. Xu, Y. Zong, X. Yi, E. Chen, **Y. Zhang**. *Executing SQL queries over encrypted character strings in the Database-As-Service model*. Knowl.-Based Syst. 35: 332-348 (2012)
41. X. Sun, H. Wang, J. Li, **Y. Zhang**. *Injecting purpose and trust into data anonymisation*. Computers & Security 30(5): 332-345 (2011)
42. A. Li, Y. Shi, J. He, **Y. Zhang**. *A Fuzzy Linear Programming-Based Classification Method*. International Journal of Information Technology and Decision Making 10(6): 1161-1174 (2011)
43. J. Ma, J. Cao, **Y. Zhang**. *Efficiently Supporting Secure and Reliable Collaboration in Scientific Workflows*, Journal of Computer and System Sciences, Vol. 76, No. 6, pages 475-489, 2010. (Q1, IF: 1.631).
44. *J. He, **Y. Zhang**, Y. Shi, G. Huang. *Domain-Driven Classification Based on Multiple Criteria and Multiple Constraint-Level Programming for Intelligent Credit Scoring*. IEEE Transaction on Knowledge and Data Engineering, Vol. 22, No. 6, pages 826-838, 2010. (Q1, IF: 1.847)
45. Y. Zong, G. Xu, P. Jin, **Y. Zhang**, E. Chen. *HC_AB: A New Heuristic Clustering Algorithm based on Approximate Backbone*, Information Processing Letter, 111(7), Pages 857-863, 2011 (Q2, IF: 0.612)
46. L. Li, G. Xu, **Y. Zhang**, M. Kitsuregawa. *Random Walk based Rank Aggregation to Improving Web Search*, Knowledge-based System, Volume 24, Issue 7, Pages 943-951, 2011 (Q1, IF: 1.574)
47. G. Xu, L. Li, **Y. Zhang**, M. Kitsuregawa and X. Yi. *Modeling User Hidden Navigational Behavior for Web Recommendation*, Web Intelligence and Agent Systems: An International Journal, Vol.9, No.3, 2011
48. Y. Hao, **Y. Zhang** and J. Cao. *Web services discovery and rank: An information retrieval approach*, Future Generation Computer Systems, Vol. 26, No. 8, pages 1053-1062, 2010. (Q1, IF: 2.365)
49. Y. Zong, G. Xu, **Y. Zhang**, H. Jiang and M. Li. *A robust iterative refinement clustering algorithm with smoothing search space*, Knowl.-Based Syst., Vol. 23, No. 5, pages 389-396, 2010. (Q1, IF: 1.574)

50. *J. He, **Y. Zhang**, G. Huang and C. Pang. *A Novel Time Computation Model Based on Algorithm Complexity for Data Intensive Scientific Workflow Design and Scheduling*, Concurrency and Computation: Practice and Experience, Vol. 21, No. 16, pages 2070-2083, 2009. (Q1, IF: 0.907)
51. **Y. Zhang** and G. Xu. *On Web Communities Mining and Recommendation*. Concurrency and Computation: Practice and Experience, Vol. 21, No. 5, pp. 561-582, 2009. (Q1, IF: 0.907)
52. X. Yi, **Y. Zhang**, and Z. Xu. *Privacy-Preserving Naive Bayes Classification on Distributed Data via Semi-Trusted Mixers*, *Information Systems*, Vol 34, No. 3, pages 371-380, 2009. (Q1, IF: 1.595)
53. *B. Zhang, **Y. Zhang** and R. Begg. *Gait Classification in Children with Cerebral Palsy by Bayesian Approach*, Pattern Recognition, Volume 42, Issue. 4, Pages 581-586, April 2009. (Q1, IF: 2.607)
54. H. Wang, **Y. Zhang** and J. Cao. *Effective Collaboration with Information Sharing in Virtual Universities*, IEEE Trans. Knowl. Data Eng., Vol. 21, No. 6, pages 840-853, 2009. (Q1, IF: 1.847)
55. H. Wang, J. Cao and **Y. Zhang**. *Delegating revocations and authorizations in collaborative business environments*, *Information Systems Frontiers*, Vol. 11, No. 3, pages 293-305, 2008. (Q1, IF: 1.596)
56. H. Wang, **Y. Zhang** and J. Cao. *Access Control Management for Ubiquitous Computing*, Future Generation Computer Systems, Vol 24, No. 8, pages 870-878, 2008. (Q1, IF: 2.365)
57. X. Yi and **Y. Zhang**. *Privacy-Preserving Distributed Association Rule Mining via Semi-Trusted Mixer*, Data and Knowledge Engineering, Vol. 63, Issue 2, pages 550-567, 2007. (Q1, IF: 1.717).
58. H. Wang, J. Cao and **Y. Zhang**. *A Flexible Payment Scheme and Its Role-based Access Control*, IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE), Vol. 17, No. 3, Pages 425-436, March 2005. (Q1, IF: 2.07).

REFEREED CONFERENCE PAPERS:

59. *S. Subramani, H. Wang, **Y. Zhang,** etc. *Mining Actionable Knowledge Using Reordering Based Diversified Actionable Decision Trees*. WISE (1) 2016: 553-560.
60. L. Sun, J. Ma, **Y. Zhang**, H. Wang. *Exploring Data Mining Techniques in Medical Data Streams.* ADC 2016: 321-332.
61. P. Vimalachandran, H. Wang, **Y. Zhang**, V. Wade and F. Whittaker. *Ensuring Data Integrity in Electronic Health Records: A Quality Health Care Implication.* Accepted by The 2016 IEEE International Conference on Orange Technologies (ICOT 2016), Australia.
62. *Supriya, Siuly, H. Wang, G. Zhuo, **Y. Zhang**. *Analyzing EEG Signal Data for Detection of Epileptic Seizure: Introducing Weight on Visibility Graph with Complex Network Feature*. ADC 2016: 56-66.
63. X. Zhou, L. Cao, L. Chen, **Y Zhang**, G. Huang, Online Video Recommendation in Social Community, Proc. Of 2015 ACM SIGMOD Conference, Melbourne, VIC, Australia, May 31 - June 4, 2015.
64. *W. Yao, J. He, H. Wang and **Y. Zhang**: Collaborative Topic Ranking: Leveraging Item Meta-data for Sparsity Reduction, *Proc. Of 29th AAAI Conference on Artificial Intelligence (AAAI-15)*: 456-464, 2015.
65. *M. Peng, J. Zhu, X. Li, J. Huang, H. Wang, and **Y. Zhang**. Central Topic Model for Event-oriented Topics Mining in Microblog Stream. CIKM '15, 1611-1620.
66. W. Yao, J. He, G. Huang, **Y. Zhang**: Modeling dual role preferences for trust-aware recommendation. *SIGIR 2014*: 975-978 (2014)
67. *W. Yao, J. He, G. Huang, **Y. Zhang**: SoRank: incorporating social information into learning to rank models for recommendation, *WWW (Companion Volume)*: 409-410 (2014).
68. H. Zheng, J. He, G. Huang, **Y. Zhang**: Optimized fuzzy association rule mining for quantitative data. FUZZ-IEEE: 396-403 (2014).
69. X. Tao, Y. Miao, **Y. Zhang**, Z. Shen: Collaborative medical diagnosis through Fuzzy Petri Net based agent argumentation. FUZZ-IEEE: 1197-1204 (2014).
70. D. Hong, **Y. Zhang**: An Exploration of Knowledge Integration: A Comprehensive View of Media Characteristics and Integration Capability. *PACIS*: 170 (2014).
71. X. Yin, B. Ng, J. He, **Y. Zhang**, D. Abbott: Unsupervised Segmentation of Blood Vessels from Colour Retinal Fundus Images. *HIS* 2014: 194-203
72. J. Liu, Z. Huang, H. Cheng, Y. Chen, H. Shen, **Y. Zhang**: Presenting diverse location views with real-time near-duplicate photo elimination. *ICDE 2013*: 505-516 .
73. Y. You, G. Xu, J. Cao, **Y. Zhang**, G. Huang: Leveraging Visual Features and Hierarchical Dependencies for Conference Information Extraction. *APWeb* 2013: 404-416
74. Z. Jiang, **Y. Zhang**, J. He, G. Huang, J. Yang: Mathematical-Morphology-Based Edge Detection of Retinal Vessels in Retinal Images. HIS 2013: 265-268
75. Z. Qiao, G. Huang, J. He, P. Zhang, L. Guo, J. Cao, **Y. Zhang**: Discovering Semantics from Multiple Correlated Time Series Stream. *PAKDD* (2) 2013: 509-520.
76. W. Yao, J. He, G. Huang, J. Cao, **Y. Zhang**: Personalized Recommendation on Multi-Layer Context Graph. WISE (1) 2013: 135-148.

77. Y. You, G. Huang, J. Cao, E. Chen, J. He, **Y. Zhang**, L. Hu: GEAM: A General and Event-Related Aspects Model for Twitter Event Detection. *WISE* (2) 2013: 319-332.
78. G. Xu, Y. Gu, P. Dolog and **Y. Zhang**, SemRec: A Semantic Enhancement Framework for Tag based Recommendation, in *the Proceedings of Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI-11)*, August 7 - 11, 2011, in San Francisco, California, USA.
79. Y. Zong, G. Xu, P. Jin, **Y. Zhang**, E. Chen, R. Pan: APPECT: An Approximate Backbone-Based Clustering Algorithm for Tags. *ADMA* (1) 2011: 175-189
80. *G. Huang, **Y. Zhang**, J. He and Z. Ding, Efficiently Retrieving Longest Common Route Patterns of Moving Objects by Summarizing Turning Regions, *The 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2011)*, pp. 375-386, 2011.
81. Z. Wu, G. Xu, R. Pan, **Y. Zhang**, Z. Hu, J. Lu: Leveraging Wikipedia concept and category information to enhance contextual advertising. *CIKM* 2011: 2105-2108.
82. L. Li, G. Xu, Z. Yang, **Y. Zhang**, M. Kitsuregawa: A Feature-Free Flexible Approach to Topical Classification of Web Queries. *SKG* 2011: 59-66
83. G. Xu, Y. Gu, **Y. Zhang**, Z. Yang, M. Kitsuregawa: TOAST: A Topic-Oriented Tag-Based Recommender System. WISE 2011: 158-171
84. *M. E. Kabir, H. Wang and **Y. Zhang**, A Pairwise-Systematic Microaggregation for Statistical Disclosure Control, *ICDM 2010*, Dec. 14-17, Sydney, Australia, pp. 266-273, 2010.
85. Y. Zong, G. Xu, P. Dolog, **Y. Zhang** and R. Liu: Co-clustering for Weblogs in Semantic Space. *WISE 2010*, pages 120-127, 2010.
86. M. Li, H. Wang and **Y. Zhang**, Optimal Privacy-aware Path in Hippocratic Databases, *DASFAA 2009:* 441-455, 2009.
87. J. Zhang, H. Shi and **Y. Zhang**, Self-Organizing Map Methodology and Google Maps Services for Geographical Epidemiology Mapping, *DICTA 2009*, pages 229-235, 2009.
88. L. Li, G. Xu and **Y. Zhang**, Masaru Kitsuregawa: Enhancing Web Search by Aggregating Results of Related Web Queries, *WISE 2009*, pages 203-217, 2009.
89. J. Ma, **Y. Zhang** and J. He, Web Service Discovery based on Latent Semantic Approach, *Proc of 2008 International Conference on Web Services*, pages 740-747, 2008.

ARC grants

Chief Investigators	**Project Tittle**	**Amount**	**Project ID**	**Period and publications**
Y. Zhang, H. Wang, X. Yi, J. Soar, Y. Miao,	Privacy Preserving Data Sharing in Electronic Health Environment	$295,000	LP150100673	2016-2018 [1, 60, 62]
Y. Zhang, G. Huang	Deep Data Mining for Anomaly Prediction from Sensor Data Streams	$329,027	DP140100841	2014-2016 [10, 11, 63-66]
Y. Zhang, J. He, P. de Souza	Using Data Mining Methods to Remove Uncertainties in Sensor Data Streams	$270,000	DP130101327	2013-2015 [12-14, 72, 76-79]
Y. Zhang, X. Yi, J. He, M. Steyn, K. Taraporewalla, J. Cao	Real-time and Self- Adaptive Stream Data Analyser for Intensive Care Management	$465,000	LP100200682	2012-2014 [15-18, 79-83]
Y. Miao and **Y. Zhang**	Data Exchange and Service Integration with Applications in Health Information Systems	$432,453	LP100100624	2010-2012 [31-33,35-37]
Zhou, Hunter, **Zhang**, Sadiq and Abal	Data Enhancement, Integration and Access Services for Smarter, Collaborative and Adaptive Whole of Water Cycle Management	$360,643	LP0882957	2008-2010 [51-54]
X. Yi and **Y. Zhang**	Privacy Protection in Distributed Data Mining	$165,000	DP0770479	2007-2009 [55, 57]
Y. Zhang, C. Liu, J. Yang and M. Papazoglou,	A Framework for Consistent Collaborative Business Transactions	$218,000	DP0557572	2005-2007 [38-39, 42, 44, 56]
Orlowska, Zhou, ... **Zhang**, ...	ARC Research Network in Enterprise Information Infrastructure	$3,200,000	ARC Research Network (RN0459895)	2004-2009 [50-53, 55, 57]

F14. Research Opportunity and Performance Evidence (ROPE) - Ten career-best research outputs

(Please upload a PDF with a list of your ten career-best research outputs, with a brief paragraph for each research output explaining its significance (five pages maximum).)

Uploaded PDF file follows on next page.

Ten career-best publications – Professor Yanchun Zhang

SIGMOD is a top conference in CS, and all other 9 papers are in tier A* journals in ARC 2010 ERA Journal Ranking and Q1 journals in Scimago Journal Rank www.scimagojr.com/.

1. X. Zhou, L. Chen, **Y. Zhang**, L. Cao, G. Huang, C. Wang. *Online Video Recommendation in Sharing Community*. SIGMOD Conference 2015: 1645-1656 (Supported by ARC DP140100841)

   ERA A, a top conference in CS (#1 in Database area).
   This paper proposed a video sharing recommendation approach based on the content and social information of videos.

2. X. Yi, **Y. Zhang**. *Equally contributory privacy-preserving k-means clustering over vertically partitioned data*. Information Systems. 38(1): 97-107 (2013) (supported by DP0770479).

   ERA Ranking: A*, Impact Factor: 1.768 (Q1)
   This paper presents an equally contributory multiparty k-means clustering protocol for vertically partitioned data, in which each party equally contributes to k-means clustering.

3. *Supriya, Siuly, H. Wang, J. Cao and **Y. Zhang**. *Weighted visibility graph with complex network features in the detection of epilepsy*. IEEE Access, Vol. 4, pp: 6554 - 6566, 2016.

   ERA Ranking: A*, Impact Factor: 1.27 (Q1)
   This paper by four CIs introduces a new idea for epilepsy detection using complex network statistical properties by measuring different strengths of the edges in natural visibility graph theory, which is considered as weight. The epileptic EEG signals are transformed into complex network and then two important statistical properties of a network such as modularity and average weighted degree used for extracting the imperative characteristics from a network of EEG signals.

4. *X. Yi, **Y. Zhang** and Z. Xu. *Privacy-Preserving Naive Bayes Classification on Distributed Data via Semi-Trusted Mixers*. Information Systems, 34(3): 371-380, 2009 (34 Citations, Supported by ARC DP0770479).

   ERA Ranking: A*, Impact Factor: 1.768 (Q1)
   This paper develops a privacy-preserving distributed association rule mining protocol based on a new semi-trusted mixer model.

5. *B. Zhang, **Y. Zhang** and R. Begg. *Gait Classification in Children with Cerebral Palsy by Bayesian Approach*. Pattern Recognition, 42(4):581-586, 2009. (27 Citations).

   ERA Ranking: A*, Impact Factor: 2.632 (Q1)
   This paper proposes a novel Bayesian classifier model with a comprehensive performances comparison for gait pattern classification.

6. *C. Sun, X. Jia, **Y. Zhang**, Y. Yang and D. Chen. *Achieving Convergence, Causality preservation and Intention Preservation in Real-time Cooperative Editing Systems*. ACM Transactions On Computer-Human Interaction, 5(1): 63-108, 1998 (**710 citations**).

   ERA Ranking: A*, Impact Factor: 1.179 (Q1)
   In this paper, a consistency model, with properties of convergence, causality preservation, and intention preservation, is proposed as a framework for consistency maintenance in real-time cooperative editing systems.

7. R. Weber and **Y. Zhang**. *An Analytical Evaluation of NIAM'S Grammar for Conceptual Schema Diagrams.* Information Systems Journal. 1996 – Blackwell Synergy (**122 citations**).

   ERA Ranking: A*, Impact Factor: 1.381 (Q1)
   This is one of the early papers on ontology analysis. This paper presents an ontological evaluation of NIAM's grammar for conceptual schema diagrams, using an ontological model proposed by Bunge, Wand and Weber.

8. **Y. Zhang** and M. E. Orlowska. *Parallel Processing for the Full Reduction of a Chain Query in Distributed Databases.* Information Systems, 18(3): 181- 187, 1993.

   ERA Ranking: A*, Impact Factor: 1.768 (Q1)
   This paper presents an efficient parallel algorithm for the full reduction of chain queries, significantly improves the query processing in distributed databases.

9. **Y. Zhang** and M. E. Orlowska. *A New Polynomial Time Algorithm for BCNF Relational Database Design. Information Systems*, 17(2):185-193, 1992.

   ERA Ranking: A*, Impact Factor: 1.768 (Q1)
   This paper formalizes the split ability of fact types in NAIM conceptual schema by using functional dependencies and develops a polynomial time algorithm for the design of BCNF relational database.

10. **Y. Zhang** and M. Orlowska. *An Improvement on the Automatic Tool for Relational Database Design*. Information Systems, 15(6): 647-651, 1990.

    ERA Ranking: A*, Impact Factor: 1.768 (Q1)
    This paper corrects some misconceptions regarding the automatic tools for relational database design. A new algorithm SYNTHESIZER+ is proposed to produce a third normal form (3NF) relational database schema with a minimum number of relations for a given set of functional dependencies (FDs).

F15. Research Opportunity and Performance Evidence (ROPE) - Further evidence in relation to research impact and contributions to the field, including those most relevant to this Proposal

(Write a maximum of 11250 characters (approximately 1500 words). Detail further evidence in relation to research impact and contributions to the field. Click on the information icon and refer to the Instructions to Applicants for the required content and formatting.)

(1) Word leading research and publication impacts

Prof. Yanchun Zhang is a world-leading expert in e-health informatics, data mining, information systems, Web & Internet, e-Environment and sensor networks fields. He has published over 280 refereed research papers in these fields, which are widely cited (over 10000 citations according to Google Scholars in December 2016) and his H-index is 36. Recent representative research outcomes are published on the following top journals and conferences related to data mining, information systems and sensor networks:

•Pattern Recognition (ERA A*, Impact Factor: 2.632)
•IEEE Transactions on Knowledge and Data Engineering (ERA A, Impact Factor: 1.892)
•Information Systems (ERA A*, Impact Factor: 1.768)
•Information Systems Journal (ERA A, Impact Factor: 1.381)
•Journal of Computer and System Sciences (ERA A*, Impact Factor: 1.0)
•The Computer Journal (ERA A*, Impact Factor: 0.755)
•IEEE Transactions on System, Man and Cybernetics-A (ERA A, Impact Factor: 2.183)
•ACM Transactions on Computer and Human Interaction (ERA A*, Impact Factor: 1.179)
•ACM Transactions on Internet Technology (ERA A, Impact Factor: 0.792)
•Future Generation Computer Systems (ERA A, Impact Factor: 1.864)
•International Conference on Data Engineering (ICDE)'13 (acceptance rate: 92/460 = 20%)
•IEEE International Conference on Data Mining (ICDM)'10, (acceptance rate: 155/793=19.6%)
•International World Wide Web Conference (WWW)'07, (acceptance rate: 111/750=14.8%)
•National Conference of the American Association for Artificial Intelligence (AAAI)'11, (acceptance rate: 242/975 = 24.8%)
•Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)'11, (acceptance rate: 90/330=27.2%)

Based on previous research in information technology, starting (since 1990) from database, then (since 2003) to Web & Internet technology, and finally (since 2008) to data mining, Prof. Zhang's resent research especially focuses on applying data mining, information system and sensor networks techniques to aid efficient knowledge discovery from environmental and health care sensor data and to promote real world applications. The recent publications on above top journals and particularly top data mining related conferences have evidenced his research now focuses on data mining and data processing fields.

(2) World class leadership and project grants

In the last 10 years, his endeavor in science research is evidenced in his related publications (for example, 38 Q1 Journal papers within last 10 years) and fames in academic circle (he is Editor-In-Chief of both World Wide Web (WWWJ) and the Journal on Health Information Science & Systems (HISS), Chair of WISE conference, and a member of Australian Research Council (ARC) College of Experts from 2008 to 2010). He also has made strong impacts through his leadership in applied research and collaborations with other universities, and industry organizations such as UQ, SEQ Healthy Waterways Partnership (HWP) and Royal Brisbane and Women's Hospital. Prof. Zhang's research has attracted 10 national level grants in data mining, database and multidisciplinary e-research fields since 2003, including 9 competitive ARC grants and 1 Australia Government grant:

• Y. Zhang, G. Xu, E. Chen and Y. Zong, "Collaboration with the University of Science and Technology of China in the field of spatiotemporal identification of health events and complex event processing", the Group Missions Component of the Australia-China Science and Research Fund (ACSRF), Department of Industry Innovation, Science, Research and Tertiary Education, Australia Government, $44,621AUD, 2013.1-2013.7.

(3) Early career researcher mentoring and project outcomes

With financial support from the above national and VU projects, he has cultivated several early career researchers and produced high quality research outcomes. An excellent example is Dr Jing He, who has achieved a DECRA project and a VU Vice-Chancellor's Peak Award. Another good example is Dr. Guangyan Huang who was also awarded a DECRA in Nov. 2013.

(4) Other awards, honors and professional services

Meanwhile, he is the winner of VU Vice Chancellor's Peak Award for Research and Research Training (Research Supervision) in 2011 and 2005 Victoria University Vice-Chancellor's Medal for Excellence in Research. He is Australia representative of International Federation of Information Processing (IFIP) Working Group 6.4 on Internet Applications Engineering, and has been the steering committee chair since 2002. He is Australian representative of IFIP Working Group 6.11: Communication Aspects of the E-World (since 2011). He is also the Editor of Web Information Systems Engineering and Internet Technology Book Series from Springer, the Chairman of International Web Information Systems Engineering Society (WISE Society). For several premium / top international conferences, he has served as a key organizer, especially for the ERA A conferences, such as vice program chair of ICDM'10, tutorial Chair of ICDE'06, as well as PC member of ICDE'07, ER'06 and ER'07. In addition, he has won outstanding research honors and awards overseas. In Japan, he was a distinguished visiting professor at Nagoya University in 2007. In China, he is currently a member of communication expert panel for Cheung Kong Scholars Programme from Ministry of Education of China (similar to ARC Federation Program) and "Thousand Talents Plan" Expert (distinguished visiting professor/expert).

F16. Currently held ARC Projects

(This information is auto-populated from your RMS profile and will include any 'active' Project which has not yet had a Final Report approved and the Project file closed by the ARC. If you have any concerns with the information recorded here, contact your Administering Organisation's Research Office. NOTE: If you hold a CI or a PI role on the Project/s listed in the table below you must ensure a progress statement is provided in G2. This requirement applies to the following schemes: Discovery Projects, Discovery Indigenous Researchers Development, Discovery Indigenous, Discovery Early Career Researcher Award, Linkage Projects, Industrial Transformation Research Hubs, Industrial Transformation Training Centres or any ARC Fellowship scheme. Please click on the information icon and refer to the Instructions to Applicants for further information.)

Identifier	Scheme Name	Investigators	Admin Organisation	Project Title	Funding	End Date
DP130101327	DP 2013 R1	Prof Yanchun Zhang ; Prof Jing He ; Prof Paulo de Souza	Victoria University	Using data mining methods to remove uncertainties in sensor data streams	$270,000	31/12/2017
DP140100841	DP 2014 R1	Prof Yanchun Zhang ; Dr Guangyan Huang	Victoria University	Deep Data Mining for Anomaly Prediction from Sensor Data Streams	$329,027	31/12/2016
LP100100624	LP 2010 R1	Prof Yuan Miao ; Prof Yanchun Zhang	Victoria University	Data Exchange and Service Integration with Applications in Health Information Systems	$325,787	30/06/2015
LP100200682	LP 2010 R2	Prof Yanchun Zhang ; Prof Xun Yi ; Prof Jing He ; A/Prof Michael Steyn ; Dr Kersi Taraporewalla ; Prof Dr Jie Cao	Victoria University	Real-time and self-adaptive stream data analyser for intensive care management	$358,334	30/06/2016
LP150100673	LP 2015 R1	Prof Yanchun Zhang ; Prof Hua Wang ; Prof Xun Yi ; Prof Jeffrey Soar ; Prof Yuan Miao ; Mr Ben Heyward	Victoria University	Privacy Preserving Data Sharing in Electronic Health Environment	$295,467	14/06/2019

Part F - Personnel and ROPE (Dr Jinli Cao)

F1. Personal Details

(To update personal details, click the 'Manage Personal Details' link below. Note this will open a new browser tab. When returning to the form ensure you 'Refresh' the page to capture the changes made to your profile.)

Participation Type

Chief Investigator

Title

Dr

First Name

Jinli

Second Name

Family Name

Cao

F2. Will you be residing predominately in Australia for the duration of the Project?

(This is a 'Yes' or 'No' question. Indicate whether you will be residing predominantly in Australia for the duration of the Project, taking into account any international travel. If you are applying as a CI and you answer 'No' to this question you will be prompted to contact your Research Office to check your eligibility.)

No

F3. Qualifications

(To update any qualifications, click on the 'Manage Qualifications' link below. Note this will open a new browser tab. When returning to the form ensure you 'Refresh' the page to capture the changes made to your profile.)

Conferral Date	AQF Level	Degree/Award Title	Discipline/Field	Awarding Organisation	Country of Award
17/10/1997	Doctoral Degree	PhD	Computer Science and Information Technology	University of Southern Queensland	Australia

F4. Are you currently undertaking a Higher Degree by Research which will be conferred after 1 January 2018?

(This is a 'Yes' or 'No' question. If you are applying as a CI and your answer is 'Yes' to this question you will be prompted to contact your Research Office.)

No

F5. Research Opportunity and Performance Evidence (ROPE) – Current and previous appointment(s) / position(s) - during the past 10 years

(To update any details in this table, click on the 'Manage Employment Details' link below. Note this will open a new browser tab. When returning to the form ensure you 'Refresh' the page to capture the changes made to your profile. Click on the information icon above and refer to the Instructions to Applicants for more information.)

Description	Department	Contract	Employment	Start Date	End	Organisation

		Type	Type		Date	
Senior Lecturer	Computer Science & Computer Engineering	Permanent	Full Time	03/01/2003		La Trobe University

F6. Employment Details as at Commencement date of Project

(This question will be used to determine your eligibility. Confirm your employment status at all organisations that you will be associated with as at the Commencement Date for the Project (1 January 2018). Enter the relevant appointment type and Full-Time Equivalent (FTE) for each organisation. Click on the information icon for further information.)

Org name	Is this an Eligible Organisation?	Please choose your appointment type for this organisation.	Please enter your FTE for this Organisation
La Trobe University	Yes	Employee	1

F7. Further Details Regarding Partner Investigator Status - Do you hold a remunerated appointment at an Eligible Organisation?

(NOTE: this question is mandatory ONLY FOR PIs WHO:
• at F2 confirmed that they will reside predominantly in Australia for the duration of the proposed Project; AND
• at F4 confirmed that they are not currently undertaking a Higher Degree by Research which will be conferred after 1 January 2018; AND
• at F6 indicated that they would hold either:
- an appointment at an Eligible Organisation equal or greater than 0.2FTE; OR
- Emeritus appointment at an Eligible Organisation

This is a 'Yes' or 'No' question. If you select 'Yes', you will be further prompted to justify your participation on this Proposal as a PI with reference to sections F7.2 and F7.3 of the Funding Rules. Click on the information icon for further information.)
Do you hold a remunerated appointment at an Eligible Organisation?

Justification of PI status

F8. Relevant Organisation for this Proposal

(Enter the Organisation that is relevant to your participation on this proposal, and that you will be associated with as at the Commencement Date for the Project (1 January 2018). The 'relevant organisation' is the primary organisation that will be supporting your involvement in this Project if it is funded. Note that the Organisation must be listed in F6 for this question to validate.)

Relevant Organisation

La Trobe University

F9. What is your time commitment to this Project?

(Enter your time commitment to this Project as a Full-Time Equivalent (FTE). Note that a FTE of 1.0 represents a full-time commitment (i.e. 5 days per week).)

0.2

F10. Are you applying for Teaching Relief?

(This is a 'Yes' or 'No' question.

• If you answer 'Yes' to this question a budget line will be automatically populated for the Teaching Relief in the budget table in Form Part D: Project Cost. This will allow you to enter the funding amount requested in the relevant year/s. To remove the Teaching Relief from the budget table you must return to this question and answer 'No'.

• Note: CIs may request funding for teaching relief or other duties in order to maximise the opportunity for the CI to conduct research. This question is only relevant for CIs and will not be activated for PIs. If the answer to this question is 'yes', the budget table in 'Project Costs' will automatically update with a line to add costs for Teaching Relief.)

No

F11. Are you applying for a Discovery International Award?

(This is a 'Yes' or 'No' question. If you answer 'Yes' to this question a budget line will be automatically populated for the DIA in the budget table in Form Part D: Project Cost. This will allow you to enter the funding amount requested in the relevant year/s. To remove the DIA from the budget table you must return to this question and answer 'No'. Click on the information icon for further information.)

No

F12. Research Opportunity and Performance Evidence (ROPE) - Details of your career and opportunities for research

(Write a maximum of 5250 characters (approximately 750 words). Please detail your career and opportunities.)

Jinli Cao obtained the PhD degree in computer science in Oct 1997 at University of Southern Queensland (USQ), Australia. Dr Cao has been a full time senior lecturer in La Trobe University (LTU) since Jan 2003. She had obtained an Early Career Researcher grant from USQ and four faculty research grants from La Trobe University. She has been given opportunities to work with other academic staff from different universities in Australia and Overseas.

Dr Cao is taking all responsibilities of academic duties for teaching and research with some administrative duties. She was a seminar coordinator in the Department of Computer Science of LTU. Dr Cao takes roughly 50% of workload on teaching, 5% of workload on administration and 45% of workload on research activities including PhD supervision and research community services. She has successfully supervised 6 PhD students and over 30 Honours & Master students in past 10 years.

Dr Cao has no career interruptions. The research mentoring and research facilities are available and supportive to Dr Cao. She has been granted 4 Faculty Block Research Grants to support some of active research projects in LTU. Dr Cao had been granted the Overseas Study Program by La Trobe University in 2009 that gave her the opportunities to collaborate with other professors in Hong Kong and China for research collaborations. Dr Cao has published 62 research papers since 2005. She is striving to get great support on the proposed research projects from ARC over the last 10 years.

Dr Cao has continuously worked with other CIs Zhang, Wang and Siuly in EEG Signals, data mining and database systems research. Scholar papers have been published as outcomes of the collaborative work between the CIs. For example:

• Supriya, Siuly, H. Wang, J. Cao and Y. Zhang. Weighted visibility graph with complex network features in the detection of epilepsy. IEEE Access, DOI: 10.1109/ACCESS.2016.2612242, 2016. (Q1; IF=1.270)

• J. Huang, M. Peng, H. Wang, J. Cao, G. Wang, X. Zhang. A Probabilistic Method for Emerging Topic Tracking in Microblog Stream. World Wide Web, 20(2): 325-350, 2017. (Q1, IF: 1.47)

In summary, Dr Cao's strong capability of achieving high quality research outcomes with 20 Q1 Journal papers in the past 10 years and her extensive research program on data mining and EEG signals means that this proposed project is highly feasible. Her collaborative research with the team and PhD student supervision demonstrates her leadership abilities.

F13. Research Opportunity and Performance Evidence (ROPE) - Significant research outputs and ARC grants

(Upload a PDF of no more than four A4 pages with a list of all research outputs, such as journal articles and refereed conference papers, book and book chapters. Use asterisks to identify research outputs relevant to this Proposal. Click on the information icon or refer to the Instructions to Applicants for the required content and formatting.)

Uploaded PDF file follows on next page.

Significant publications and ARC grants: Jinli Cao

Dr Cao has published 91 refereed scholar papers, including 24 Q1 Journal papers in data mining, database, privacy preserving and Web services. Representative publications are on *Information Sciences, ACM Transaction on Internet Technology, IEEE Transactions on Knowledge and Data Engineering, IEEE Trans on Parallel and Distributed Systems and World Wide Web.*

REFERRED JOURNAL ARTICLES:

1. *Supriya, Siuly, H. Wang, **J. Cao** and Y. Zhang. Weighted visibility graph with complex network features in the detection of epilepsy. IEEE Access, DOI: 10.1109/ACCESS.2016.2612242, 2016. (Q1; IF=1.270)
2. Y. Wang, Y. Shen, H. Wang, **J. Cao**, X. Jiang. MtMR: Ensuring MapReduce Computation Integrity with Merkle Tree-based Verifications, IEEE Transactions on Big Data, no.99, pp.1-16, 2016.
3. *J. Huang, M. Peng, H. Wang, **J. Cao**, G. Wang, X. Zhang. *A Probabilistic Method for Emerging Topic Tracking in Microblog Stream*, World Wide Web, pp: 1—26, 2016. (Q1, IF: 1.47)
4. *H. Nyuyen and **J. Cao**. *Trustworthy Answers For Top-K Queries On Uncertain Big Data In Decision Making.* Information Sciences, Volume 318, 10 October 2015, Pages 73–90. (Q1, IF: 4.22)
5. T. Le, **J. Cao**, Z. He. *Answering skyline queries on probabilistic data using the dominance of probabilistic skyline tuples*. Information Sciences, accepted by September 2015. (Q1, IF: 4.22)
6. * L. Vo**, J. Cao**, Wenny Rahayu. *Structured content-based query answers for improving information quality***.** World Wide Web, July 2015, Volume 18, Issue 4, pp 889-912. First online: 26 March 2014. (Q1, IF: 1.47)
7. K. Puansurin and **J. Cao**. *Development of Traffic Assessment Information System*. IEICE Transaction on Information and Systems. accepted by October, 2015.
8. K. Puansurin and **J. Cao**. *The Development of Product Distribution System for New Distribution Centers Using Simulation Techniques*. International Journal of Information Technology, Control and Automation, 5(3) 1-18 (2015). (Q1, IF: 0.95)
9. D. Liu, **J. Cao**. *Scheduling para-virtualized virtual machines based on events*. Future Generation Comp. Syst. 29(5): 1130-1139 (2013). (Q1, IF: 2.46)
10. *T. Le, **J, Cao**, Z. He. *Top-k best probability queries and semantics ranking properties on probabilistic databases*. Data Knowl. Eng. 88: 248-266 (2013). (Q2, IF: 1.59)
11. *L. Vo, **J. Cao**, W. Rahayu, H. Nguyen. *Structured content-aware discovery for improving XML data consistency.* Information Sciences 248: 168-190 (2013). (Q1, IF: 4.22)
12. Y. Hao, Y. Zhang, J. Cao. *A novel QoS model and computation framework in web service selection*. World Wide Web 15(5-6): 663-684 (2012).). (Q1, IF: 1.45)
13. *K Nguyen, **J. Cao**. *Top-K data source selection for keyword queries over multiple XML data sources*. Journal of Information Science 38(2): 156-175 (2012). (Q1, IF: 1.86)
14. *K. Nguyen, **J. Cao.** *Top-k Answers for XML Keyword Queries*. World Wide Web, 15(5-6): 485-515 (2012). (Q1, IF: 1.45)
15. J. He, Y. Zhang, G. Huang, **J. Cao**. *A smart web service based on the context of things*. ACM Trans. Internet Techn. 11(3): 13 (2012). (Q1, IF: 1.29, citations 32)
16. G. Huang, .Y. Zhang, J. He, **J. Cao**. *Fault Tolerance in Data Gathering Wireless Sensor Networks*. Comput. J. 54(6): 976-987 (2011).
17. S. Wang, **J. Cao**, J. Hu. *A frequency domain subspace blind channel estimation method for trailing zero OFDM systems.* J. Network and Computer Applications 34(1): 116-120 (2011). (Q1, IF: 2.22)
18. *Hao, Y. Zhang, **J. Cao.** *Web services discovery and rank: An information retrieval approach*. Future Generation Computer. Systems. 26(8): 1053-1062, 2010. (Q1, IF: 2.46, citations 49)
19. *Y. Xin, Z. He and **J. Cao**. *Effective Pruning for XML Structural Match Queries.* Journal of Data & Knowledge Engineering. 69(6): 640-659, 2010. (Q2, IF: 1.59)
20. J. Ma, **J. Cao** and Y. Zhang. *Efficiently Supporting Secure and Reliable Collaboration in Scientific Workflows*. Journal of Computer and System Sciences. 76(6), 475-489, 2010.
21. M. Tang, **J. Cao** & X. Jia. *Optimization on Merging Multiple Range-based Queries in Wireless Sensor Networks*. Accepted by *IEEE Trans on Parallel and Distributed Systems*. (Q1, IF: 2.17)
22. H. Wang, **J. Cao** and Y. Zhang. *Effective collaboration with information sharing in virtual universities*. IEEE Transactions on Knowledge and Data Engineering. 21(6), pp. 840-853, 2009. (Q1, IF: 2.067, citations 32)
23. *H. Wang, **J. Cao** and Y. Zhang. *Delegating revocations and authorizations in collaborative business environments*. *Journal of Information Systems Frontiers*, pp. 293-305, Vol. 11, Issues 3, 2009. (Q1, IF:1.08)
24. *H. Wang, Y. Zhang & **J. Cao**. *Access control management for ubiquitous computing*. The International Journal of Future Generation Computer Systems, Vol. 24, pp 870-878, 2009. (Q1, IF: 2.46, citations 39)
25. Bilal Choudry, P. Bertok and **J. Cao**. *Cost based web services transactions management*. The International Journal Web and Grid Services, Vol. 2, pp. 198-220, 2006. (Q2, IF: 1.58)
26. X. Jia, D. Li, H. Du, and **J. Cao**. *On Optimal Replication of Data Object at Hierarchical and Transparent Web Proxies.* IEEE Trans on Parallel and Distributed Systems. 16(8): 673-685,, 2005. (Q1, IF: 2.17, citations 33)
27. *H. Wang, **J. Cao**, Y. Zhang. *A flexible payment scheme and its role based access control*. IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 3, pp. 425-436, 2005. (Q1, IF: 2.067, citations 66)

28. J. Hou and **J. Cao**. *Matrix Model for Web Page Community Construction*. International Journal of Science and Research (2005). Vol. 1 (1), Page 21-30, 2005.
29. H. Wang, **J. Cao**, Y. Zhang, X. Huang. *A flexible electronic payment scheme and its management.* Information, Vol. 7, pp. 187-196, No. 2, 2004.
30. H. Wang, Y. Zhang, **J. Cao,** Y. Kambayahsi. *A global ticket-based access scheme for mobile users. Information Systems Frontiers*, Vol. 6, No 1, pp. 35-46, 2004**.** (Q1, IF: 1.08, citations 15)
31. H. Wang, **J. Cao**, Y. Zhang. *An Electronic Payment Scheme and Its RBAC management*. Concurrent Engineering: Research and Application Journal. Volume 12, Number 3, pp.247- 257, Sept. 2004. (Q1, IF: 0.85)
32. *H. Wang, Y. Zhang, **J. Cao**, V. Varadharajan. *Achieving Secure and Flexible M-Services Through Tickets*. IEEE Transactions on Systems, Man, and Cybernetics (IEEE TSMC). Part A: Systems and Humans, Vol. 33, No.6, pages: 697- 708, Nov. 2003. (Q1, IF: 6.22, citations 35)
33. H. Wang, **J. Cao**, Y. Zhang. *An Electronic Payment Scheme and Its RBAC management*. Concurrent Engineering: Research and Application Journal. Volume 12, Number 3, pp.247- 257, Sept. 2004. (Q1, IF: 0.85)
34. J. Hou, Y. Zhang, **J. Cao**, W. Lai and D. Ross. *Visual Support for Information Retrieval based on Linear Algebra.* Journal of Applied Systems Studies, Vol. 3, No.2, 2002, Cambridge International Science Publishing.
35. C. Liu, X. Zhou, **J. Cao** and X. Lin. *Incremental Support of Global Update in OzGateway - a Cooperative Database System.* Informatica & International Journal of Computing and Informatics, Vol. 24 No. 1, March, 2000, Pg. 27-31, ISBN 0350-5596.
36. X. Zhou, X. Lin, C. Liu and **J. Cao**. *Spatial Query Processing Using Generalized Filter*. International Journal of Cooperative Information Systems, Vol . 9, Nos. 1 & 2 (2000), pp. 77-91. World Scientific Publishing Company.

REFERRED CONFERENCE PAPERS:

37. *H. Nguyen, **J. Cao**. *Preference-Based Top-k Representative Skyline Queries on Uncertain Databases*. PAKDD (2) 2015: 280-292
38. K. Puansurin, **J. Cao**. *Evaluating the capability of new distribution center using simulation techniques*. Proceedings of 3rd International Conference of Information Technology, Control and Automation, pp. 43-59, 2015.
39. K. Puansurin, **J. Cao**. *Simulation application for improving the efficiency of new distribution center*. Proceedings of 19th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications DS-RT 2015.
40. K. Puansurin, **J. Cao**. *The development of product distribution system for new distribution centers using simulation techniques.* International Journal of Information Technology, Control and Automation IJITCA, vol. 5, no. 3, pp. 1-18, 2015.
41. D. Liu, **J. Cao**, and Jie. Cao. *FEAS: A full-time event aware scheduler for improving responsiveness of virtual machines.* Proc. Australasian Computer Science Conference (ACSC 2012), Melbourne, Australia. CRPIT, 122. Reynolds, M. and Thomas, B. Eds., ACS. 3-9.
42. *Y,.Yang and **J. Cao**. *Feature-based recommendation framework on OLAP*. Proc. Australasian Database Conference (ADC 2012), Melbourne, Australia. CRPIT, 124. Zhang, R. and Zhang, Y. Eds., ACS. 81-88.
43. *T. Le and **J. Cao**: Top-k Best Probability Queries on Probabilistic Data, in Proceedings of DASFAA (2) 2012: 1-16.
44. *L. Vo, **J. Cao** & W, Rahayu. *Discovering Conditional Functional Dependencies in XML Data*. Proceedings of 22nd Australasian Database Conference , 16-21- Jan, 2011
45. *K. Nguyen, **J. Cao.** *Relevant Answers for XML Keyword Search: A skyline Approach*. Proceedings of the 11th International Conference of Web Information Systems Engineering, pp: 216-224, 2010.
46. *K. Nguyen, **J. Cao.** *Exploit Keyword Query Semantics and Structure of Data for Effective XML Keyword Search*. Proceedings of *The 21st Australasian Database Conference*, pp133-140**,** 2010.
47. B. Choudry, **J. Cao**. *Coordinating Business Web Services*. Proceedings of 12th International Conference on Information Integration and Web-based Applications & Services (iiWAS2010), pp 89-95, France, 2010.
48. M. Tang, **J. Cao**, X. Jia & K. Liu. *Optimization on Data Object Compression and replication in Wireless Sensor Networks*. Proceedings of Database Systems for Advanced Applications (DASFAA), LNCS 5463, 2009, pp. 77-91. Brisbane, Australia.
49. Y. Hao, **J. Cao** & Y. Zhang. *Efficient IR-Style Search over Web Services*. Proc. of The 21st International Conference on Advanced Information Systems Engineering (CAiSE'09), pages 305-418, Volume 5565/2009 LNCS, 8-12 June 2009, Amsterdam, The Netherlands.
50. H. Shen, Z. Huang, **J. Cao** and X. Zhou. *High-dimensional indexing with oriented cluster representation for multimedia database*. Proc. of IEEE International Conference on Multimedia and Expo, pp. 1628 – 1631, 2009. (Invited)

51. M. Tang, **J. Cao** & X. Jia. *Efficient Power Management for Wireless Sensor Networks: a Data-Driven Approach*. Proceedings of the 33rd Annual IEEE Conference on Local Computer Networks (LCN), pp. 106-113, 2008.

52. M. Tang & **J. Cao**. *An Energy-Efficient Data-Driven Power Management for Wireless Sensor Networks*. Proceedings of 5th International Workshop on Data Management for Sensor Networks. *VLDB 2008*, pp. 21-27.

53. Duy-Bo Dao, **J. Cao.** *A Glance on Current XML Twig Pattern Matching Algorithms*. Proceedings of International Conference of Computational Science and Its Applications - *ICCSA 2008*, Perugia, Italy, pp 307-321 LNCS 5073, 2008

54. *K. Nguyen, **J. Cao** and C. Liu. *Semantic-Enabled Organisation of Web Services*. 10th Asia-Pacific Conference, APWeb 2008, Shenyang, China, vol. LNCS 4976, pp. 511-521.

55. Le and **J. Cao**. *Flexible and Semantics-Based Support for Web Services Transaction Protocols*. Proceedings of The 3rd International Conference on Grid and Pervasive Computing), 2008, vol. LNCS 5036, pp. 492-503.

56. *Y. Hao, Y. Zhang and **J. Cao.** *WSXplorer: Searching for Desired Web Services.* Proceedings of the 19th International Conference on Advanced Information Systems Engineering (CAiSE'07), pp. 173-187, 11-15 June 2007, Trondheim, Norway.

57. J. Ma, **J. Cao** and Y. Zhang. *Probabilistic Semantics Approach for Discovering Web Services*. *Proceedings of the 16th International Conference on World Wide Web (WWW 2007),* May 8-12, 2007, Banff, Canada.

58. *Ling Wang and **J. Cao**. *Web Services Semantic Searching enhanced by Case Reasoning*. Proceedings of 18th International workshops on Database and Expert Systems Applications, 3-7 Sep 2007, Regensburg, Germany. The 3rd Ubiquitous Web Systems and Intelligence Workshop (UWSI 2007), pp.565-569.

59. *H. Wang and **J. Cao**. *Delegating revocations and authorizations*. The International Workshop on Collaborative Business Processes (CBP 2007) in conjunction with the 5th International Conference on Business Process Management (BPM 2007), pp. 294-305**,** 24-27 September 2007, Brisbane, Australia.

60. B. Choudry and **J. Cao**. *Towards Dynamic Integration of Heterogeneous WEB Services Transaction Partners*. Proceedings of 9th International Conference on Information Integration and Web-based Applications & Services (iiWAS2007).

61. M. Tang and **J. Cao**. Optimization on Distributed User Management in Wireless Sensor Networks", *Proceedings of 13th International Conference on Parallel and Distributed Systems (ICPADS)* 2007. pp. 1-8, December 5 - 7, 2007, Taiwan.

62. M. Tang, **J. Cao** and N. Chilamkurti. *TAMPA: Tabu Search-based Multiple Queries Optimization for Wireless Sensor Networks*. Proceedings of the Third International Conference on Wireless Communications, Networking and Mobile computing (WiCOM2007), pp. 2731-2734, Sep,. 21-23, 2007. ShangHai China.

63. M. Tang and **J. Cao**. *A dynamic mechanism for handling mobile computing environmental changes*. *Proceedings of the First International Conference on Scalable Information Systems* (INFOSCALE 2006), ACM Society, 29 May - 02 June, 2006, Hong Kong. pp. 1-9.

64. *H. Wang, Y. Zhang and **J. Cao** (2006). *Ubiquitous Computing Environments and Its Usage Access Control*. *The First International Conference on Scalable Information Systems,* InfoScale *2006*, Hong Kong, pp. 1-9.

65. *H. Wang, **J. Cao** and L. Sun**.** *Role-based delegation with negative authorization*. The Eighth Asia Pacific Web Conference (APWEB06), LNCS, pp. 307-318, Harbin, China, 2006.

66. R. Logothetis and **J. Cao**. *Web Services Oriented Transactions Using Partial dependency*. IEEE ITI 3rd International Conference on Information & Communications Technology, pp. 619-632, 2005.

67. B. Choudry, P. Bertok and **J. Cao**. *A cost based coordination model for long running transactions in Web services.* Proceedings of the Seventh International Conference on Information Integration and Web Based Applications & Services (IIWAS2005), page 97-108, 19-21 September 2005, in Kuala Lumpur, Malaysia.

68. **J. Cao**, L. Sun & H. Wang. *Towards secure XML document with usage control*. The Seventh Asia Pacific Web Conference (APWEB05), LNCS 3399, pp. 296-307, ShangHai, China, March29-April 2, 2005.

69. *H. Wang, L. Sun, Y. Zhang, **J. Cao**. *Authorization Algorithms for the Mobility of User-Role Relationship*. Proceedings of the 28th Australasian Computer Science Conference-2005, pp. 69-78, Newcastle, NSW, Australia, January 31 - February 3, 2005

70. *H. Wang, **J. Cao** and Y. Zhang. *Identifying Role-Based Access Constraints With Object Constraint Language.* Proceedings of Sixth Asia Pacific Web Conference (APWEB 2004), Lecture Notes in Computer Science, pp. 687-696. Hangzhou, China, April 14-17, 2004.

71. H. Wang, L. Sun, Y. Zhang and **J. Cao**. *Anonymous access scheme for electronic services*. Proceedings of the Twenty-Seventh Australasian Computer Science Conference(ACSC2004), Australian Computer Society, Vol. 26, pp, 295-304. Dunedin, New Zealand, 2004.

72. J. Hou, Y. Zhang, **J. Cao**. *Web Page Clustering: A Hyperlink-Based Similarity and Matrix-Based Hierarchical Algorithms*. Proceedings of 5th Asian-Pacific Web Conference (APWeb 2003), LNCS2642, pp. 201-212.

73. H. Wang , **J. Cao** and Y. Zhang. *Formal authorization approaches for permission-role assignment*. Proceedings of 15th Australasian Database Conferences (ADC2003), pp 125-133, 2003.

74. *H. Wang, Y. Zhang and **J. Cao**. *A Flexible Payment Scheme and its Permission-Role Assignment*. Proceedings of Twenty-Sixth Australasian Computer Science Conference, pp. 189-198, 2003

75. J. Hou, Y. Zhang and **J. Cao**. *Web Page Clustering: A Hyperlink-Based Similarity and Matrix-Based Hierarchical Algorithms, , Web Technologies and Applications*, The 5th Asian-Pacific Web Conference, APWeb 2003, Proceedings. Lecture Notes in Computer Science 2642, pp: 201-212.

76. *H. Wang, **J. Cao**, Y. Zhang. *A Flexible Payment Scheme and Its User-role Assignment*, Cooperative Internet Computing, Editor Alvin Chan, published by Kluwer Academic Publisher, pp. Dec. 2002.

77. H. Wang, **J. Cao** & Y.Kambayashi. *Building Consumer Self-anonymity Scalable Payment Protocol for Internet Purchases.* Proceedings of 12th International Workshop on Research Issues on Data Engineering: Engineering E-Commerce/E-Business Systems (RIDE-2EC'2002), pp. 159-168, San Jose, USA.

78. *H. Wang, **J. Cao**, Y. Zhang. *Formal Authorization Allocation Approaches for Role-Based Access Control Based on Relational Algebra Operations*, The 3nd International Conference on Web Information Systems Engineering (WISE'2002), pp. 301-310, Dec. 3-6, 2002, Singapore.

79. H. Wang, **J. Cao**, Y. Zhang. *A Flexible Payment Scheme and Its Role-based User-role Assignment*, Second International Workshop on Cooperative Internet Computing, pp. 58-68, 2002, Hong Kong.

80. **J. Cao,** L. Wang and X. Jia. *Minimize Execution Time of Distributed Transactions on Heterogeneous Databases.* Proceedings of the IASTED International Conference on Networks, Parallel and Distributed Processing and Applications (NPDPA02), pp. 229-234, Oct 1-4, 2002, Tsukuba, Japan.

81. H. Wang, **J. Cao**, Y. Zhang. *Ticket-Based Service Access Scheme for Mobile Users*. Twenty-Fifth Australian Computer Science Conference (ACSC2002), pp: 285-292, 2002.

82. **J. Cao** and Marian M. Orlowski. *On Performance of Cooperative Transactions in Multidatabase Systems*. Proceedings of the Sixteenth International Symposium on Computer and Information Sciences (ISCIS XVI), pp. 120-127, Turkey, 5-8 Nov,2001.

83. H. Wang, **J.Cao**, & Y.Zhang. *A Consumer Anonymity Scalable Payment Scheme with Role Bases Access Control*. Proceedings of The Second International Conference on Web Information Systems Engineering (WISE 2001), pp. 53-62, 3-6 December 2001, Kyoto, Japan.

84. J. Hou, Y. Zhang, **J. Cao** & W. Lai. *Visual Support for Text Information Retrieval Based on Matrix's Singular Value Decomposition*. Proceedings of The 1st International Conference on Web Information Systems Engineering (WISE2000), pp. 344-351, Hong Kong, 19-20 June 2000.

85. C. Liu, X. Zhou, **J. Cao** and X. Lin. *Global Transaction management in A Cooperative Database System Database, Web and Cooperative Systems*. Proceedings of 1st international Symposium on Database, Web and Cooperative Systems (DWACOS'99), pp.73-78, Baden-Baden, August, 1999.

86. X. Meng, S. Wang, X. Jia and **J. Cao**. *Natural Language Query Optimization and Translation in Chinese Database System.* Proceedings of 2nd International Symposium on Cooperative Database Systems for Advanced Applications (CODAS'99), pp. 340-351. Wollongong.

87. **J. Cao** and Marian M, Orlowski. *The effect Semijoin Reduction on Query Processing In Distributed Database Systems*. Proceedings of The Twelfth International Symposium On Computer And Information Sciences (ISCIS-XII), 1997, Antalya, Turkey.

88. **J. Cao** and M. Orlowski. *Transactions Classification and a concurrency Control Algorithm in a Multidatabase System.* Proceedings of "11th International Symposium on Computer and Information Systems (ISCIS-XI)" 1996, Antaly, Turkey.

89. **J. Cao** and Marian Orlowski. *On Transaction Recovery in a Multidatabase System*. Proceedings of InterSymp-96, The International Conference on System Research" in Germany, 1996.

90. **J. Cao** and Marian W. Orlowski. *On Partial Order Processing for Dynamic Concurrency Control in Database System.* Proceedings of "International Conference for Young Computer Scientists'9 5", pages 259-264, ISBN 7-301-02880-6, Peking University Press, 1995.

91. **J. Cao** and Marian W. *Orlowski. On Dynamic Concurrency Control in a Multidatabase System.* Advances in Database and Expert Systems, Volume 2, Pages 11-15, ISBN 0921836333, 1995.

92. M. Orlowski and **J. Cao**. *On Optimization of Update Propagation in Multidatabase Systems*. Proceedings of "IEEE International Conference of TENCON' 93".

SCHOLARLY BOOKS:

93. J. Cao, X. Jia and L. Shu. Multimedia Data Applications In Wireless Sensor Networks, International Journal Sensor Networks. 2012.

F14. Research Opportunity and Performance Evidence (ROPE) - Ten career-best research outputs

(Please upload a PDF with a list of your ten career-best research outputs, with a brief paragraph for each research output explaining its significance (five pages maximum).)

Uploaded PDF file follows on next page.

Ten career-best publications – Dr Jinli Cao

All 10 papers are in Q1 ranking (http://www.scimagoir.com/, Journal Rankings) and with impact factors information.

1. H. Nyuyen and **J. Cao**. Trustworthy Answers For Top-K Queries On Uncertain Big Data In Decision Making. *Information Sciences*. Vol. 318: 73–90, 2015.

Impact Factor: 4.22 (Q1). This paper presents a novel approach called Dominating Top-k Aggregate Query (DA-Topk) to provide trustworthy and reliable informative knowledge from uncertain Big Data by combining the techniques of skyline and top-k queries.

2. *J. Huang, M. Peng, H. Wang, **J. Cao**, G. Wang, X. Zhang. A Probabilistic Method for Emerging Topic Tracking in Microblog Stream, *World Wide Web*, 20(2): 325-350, 2017. (Q1, IF: 1.47)

Impact Factor: 1.76 (Q1). This paper by the three CIs' collaborations in the recent project to propose a novel emerging topics tracking method, which aligns emerging word detection from temporal perspective with coherent topic mining from spatial perspective.

3. *Supriya, Siuly, H. Wang, **J. Cao** and Y. Zhang. Weighted visibility graph with complex network features in the detection of epilepsy. *IEEE Access*, Vol. 4, pp: 6554 – 6566, 2016.

Impact Factor: 1.27 (Q1). This paper by the CIs introduces a new idea for epilepsy detection using complex network statistical properties by measuring different strengths of the edges in natural visibility graph theory, which is considered as weight. The epileptic EEG signals are transformed into complex network and then two important statistical properties of a network such as modularity and average weighted degree used for extracting the imperative characteristics from a network of EEG signals.

4. X. Jia, D. Li, H. Du, and **J. Cao**. On Optimal Replication of Data Object at Hierarchical and Transparent Web Proxies. *IEEE Trans on Parallel and Distributed Systems.* 16(8): 673-685, 2005 (citations 36).

Impact Factor: 2.17 (Q1). This paper investigated a crucial problem for optimising the data allocations with proxies having unlimited storage capacities, or limited storage capacities. Two novel heuristic algorithms are explored and proved their out-performance.

5. L. H. Vo, **J. Cao**, W. Rahayu, H. Nguyen. Structured content-aware discovery for improving XML data consistency. *Information Sciences.* Vol. 248: 168-190, 2013.

Impact Factor: 4.22 (Q1). It exploits the semantics of data structures to detect similar paths from the sources, from which a data summary is constructed as an input for the discovery process. This aims to avoid returning redundant data rules due to structural inconsistencies.

6. H. Wang, **J. Cao**, Y. Zhang. Access control management for ubiquitous computing. *Future Generation Computer Systems*. Vol. 24: 870-878, 2008. Outcomes from ARC DP0663414 (38 citations)

Impact Factor: 2.79 (Q1). This paper by the three CIs' collaborations in the recent project discusses a usage control model to protect services and devices in ubiquitous computing environments, which allows the access restrictions directly on services and object documents.

7. H. Wang, Y. Zhang, **J. Cao**. Effective collaboration with information sharing in virtual universities. *IEEE Transactions on Knowledge and Data Engineering*. 21(6): 840-853, 2009. Outcomes from ARC DP0663414 (40 citations).

Impact Factor: 2.07 (Q1). This paper with CIs Zhang and Wang presents a new rule-based framework to identify challenges of sharing in virtual university environments through role-based access control.

8. K Nguyen, **J. Cao**. Top-K data source selection for keyword queries over multiple XML data sources. *Journal of Information Science.* 38(2): 156-175, 2012.

Impact Factor: 1.86 (Q1). This paper proposes a novel approach for selecting the top-K data sources by relying on their relevance to a given query, to avoid the high cost of searching in numerous, potentially irrelevant data sources. The approach summarizes the data sources as succinct synopses for the rapid filtering of non-promising sources..

9. H. Wang, Y. Zhang, **J. Cao,** Y. Kambayahsi. A global ticket-based access scheme for mobile users. *Information Systems Frontiers*. 6(1): 35-46, 2004 (18 citations).

Impact Factor: 1.08 (Q1). This paper collaborated by the three CIs presents a ticket-based access model for mobile services. The model supports efficient authentication of users, services and service providers over different domains. Tickets are used to verify correctness of the requested service as well as to direct billing information to the appropriate user.

10. H. Wang, **J. Cao**, Y. Zhang. A flexible payment scheme and its role based access control. *IEEE Transactions on Knowledge and Data Engineering (TKDE).* 17(3): 425-436, 2005 (724citations).

Impact Factor: 2.07 (Q1). This paper with CIs Zhang and Wang proposes a practical payment protocol with scalable anonymity for Internet purchases, and analyzes its role-based access control management.

F15. Research Opportunity and Performance Evidence (ROPE) - Further evidence in relation to research impact and contributions to the field, including those most relevant to this Proposal

(Write a maximum of 11250 characters (approximately 1500 words). Detail further evidence in relation to research impact and contributions to the field. Click on the information icon and refer to the Instructions to Applicants for the required content and formatting.)

CI Cao's research areas are including Big Data analytics, Data quality, Data Reliabilities, Data security, Uncertain Big data Queries, Top-K Query Ranking, Recommendation systems and Cloud Computing that are closely related to the proposed project. She has been closely working with other CIs for a long time. For example, working with CI Wang since 2001 and have published over 29 joint research papers. The joint research has significant contributions to the fields of data analytics on EEG data and data mining algorithms and system implementation. The collaborated research outcomes have laid the proposed project and been the evidence for the success of the project.

She is investigating a new approach for capturing, storing, and analysing data in a timely and cost-effective way. The outcome of CI Cao's project can be combined with this proposed new project that the multiple associated data for brain signal processing, data compression and prediction of neurological abnormalities from massive brain signal data. The novel optimised algorithms will be devised for the computation of health monitoring in the real time. Dr Cao 's extensive research experiences and skills of optimization query processing and data management will contribute the success of the proposed project.

Dr Cao has been contributing many related research communities. She has been an ARC Reader since 2006, and an external grants accessor for Research Grants Council (RGC) of Hong Kong since 2011. She was the Guest Editor for special issue: Multimedia Data Applications In Wireless Sensor Networks, International Journal Sensor Networks in 2012. She was the Editorial Reviewer Board member for the International Journal of Distance Education Technologies, USA for 3 years. In addition, Dr Cao has served as a technical program committee member for 40+ international workshops/conferences in the past 10 years such as:
•The 25th Australasian Database Conference (ADC 2014), Brisbane, Australia, July, 2014 (PC member)
•The 24th Australasian Database Conference (ADC 2013), Adelaide, South Australia, 29 January - 1 February, 2013 (PC member)
•iiWAS 2014 : The 16th International Conference on Information Integration and Web-based Applications & Services(PC member)
•The International Conference on Data and Knowledge Engineering (ICDKE), Fujian, China, November 21-23, 2012 (PC member)
•The 14th Asia-Pacific Web Conference (APWeb), April 11-13, 2012, Kunming, China (PC member)

F16. Currently held ARC Projects

(This information is auto-populated from your RMS profile and will include any 'active' Project which has not yet had a Final Report approved and the Project file closed by the ARC. If you have any concerns with the information recorded here, contact your Administering Organisation's Research Office. NOTE: If you hold a CI or a PI role on the Project/s listed in the table below you must ensure a progress statement is provided in G2. This requirement applies to the following schemes: Discovery Projects, Discovery Indigenous Researchers Development, Discovery Indigenous, Discovery Early Career Researcher Award, Linkage Projects, Industrial Transformation Research Hubs, Industrial Transformation Training Centres or any ARC Fellowship scheme. Please click on the information icon and refer to the Instructions to Applicants for further information.)

Part G - Research Support and Statements on Progress (DP180103563)

G1. Research support for all Participants

(For each participant on this Proposal, provide details of:
i) current submitted ARC proposals (i.e. for which the outcome has not yet been announced);
ii) any newly funded ARC Projects which are not yet showing in the Participant's question (Currently held ARC Projects); and
iii) research funding from non-ARC sources (in Australia and overseas). For research funding from non-ARC sources, list all projects/proposals/awards/fellowships awarded or requests submitted involving that participant for funding for the years 2017 to 2022 inclusive.)

Uploaded PDF file follows on next page.

Current ARC Proposals and newly founded ARC Projects which are not yet active									
Description (all named investigators on any proposal or grant/ project/ fellowship in which a participant is involved, project title, scheme and round)	**Same Research Area** (Yes/No)	**Support Status** (Requested/Current/Past)	**Proposal/ Project ID**	**2017** ($'000)	**2018** ($'000)	**2019** ($'000)	**2020** ($'000)	**2021** ($'000)	**2022** ($'000)
H. Wang, Y. Zhang, J. Cao, *Increasing data quality with group associations in outsourcing environment, ARC DP18*	Yes	R	DP180103563		148	148	148		
H. Wang and X. Jiang, *Detection and prevention of data exfiltration by insiders, ARC DP18*	No	R	DP180103544		149	152	156		
Y. Zhang, X. Zhou, L. Chen, *Graph-based Context-aware Recommendation in Big Social Media Data, ARC DP18*	No	R	DP180103451		145	149	153		
Y. Zhang, H. Wang, Siuly, J. Cao, V. Wade *Analyzing neurological abnormalities from brain signal data, ARC LP16*	No	R	LP160101736	99	126	127			

Funding from non-ARC sources									
Description (all named investigators on any proposal or grant/ project/fellowship in which a participant is involved, project title, source of support, scheme and round)	**Same Research Area** (Yes/No)	**Support Status** (Requested/Current/Past)	**Proposal/ Project ID** (for NHMRC proposals only)	**2017** ($'000)	**2018** ($'000)	**2019** ($'000)	**2020** ($'000)	**2021** ($'000)	**2022** ($'000)

G2. Statements on Progress for ARC-funded Projects

(A progress statement must be provided for any currently funded ARC Project that involves a Participant on this Proposal as a CI or PI. This requirement only applies to funding held under the Discovery Projects, Discovery Indigenous Researchers Development, Discovery Indigenous, Discovery Early Career Researcher Award, Linkage Projects, Industrial Transformation Research Hubs, Industrial Transformation Training Centres or any ARC Fellowship scheme. Click on the information icon or refer to the Instructions to Applicants for further information.)

Project ID

DP130101327

First Named Investigator

Yanchun Zhang

Scheme

Discovery Projects

Statement

Uploaded PDF file follows on next page.

Project ID: DP130101327, Using Data Mining Methods to Remove Uncertainties in Sensor Data Streams

CIs: Prof. Yanchun. Zhang (Victoria University), Dr. Jing He (Victoria University), Dr. Paulo de Souza (CSIRO)

This project is to develop a flexible RuuDM (Removing uncertainties using Data Mining) framework to resolve the three uncertainties: missing temporal points, missing stream pieces and measurement errors and tackle two challenges: the uncertainties reduce the efficiency and accuracy of querying sensor data streams; and the uncertainties in sensor streams make querying multiple streams more difficult than querying single streams.

The research progress achieved over the period is outlined below:

(1) We proposed a coloring based scheduling method to schedule all groups to work in a sequence of time slots. We then simulated the clique based scheduling algorithm in a network with scheduling protocol with TDMA framing structure. The experimental results show that the proposed method is able to deal with the challenges of uncertain data streams.

(2) We designed a data streams analysis model and algorithms that not only can compress redundant data but also correct missing data. In our experimental study, we took the querying of longest common route patterns from various sizes of sensor stream datasets as an example to validate the accuracy and efficiency of the proposed method. We also summarized the original sensor streams by using inflexions, and then grouped close inflexions into the same cluster. To deduce the implicit common regions, we provided discovering implicit semantic places (DISP) procedure so that query common patterns can be directly issued on exact data sequences.

(3) We also applied fuzzy petri nets as the knowledge model to handle the uncertain data in reasoning. Uncertain rules are used when the knowledge is mined from databases.

Publications

- J. Ma, L. Sun, H. Wang, Y. Zhang, U. Aickelin. Supervised Anomaly Detection in Uncertain Pseudoperiodic Data Streams. ACM Trans. Internet Techn. 16(1): 4, 2016.
- L. Sun, J. Ma, Y. Zhang, H. Dong, F. K. Hussain. Cloud-FuSeR: Fuzzy ontology and MCDM based cloud service selection. Future Generation Comp. Syst. 57: 42-55, 2016.
- X. Zhou, J. He, G. Huang, Y. Zhang. SVD-based incremental approaches for recommender systems. J. Comput. Syst. Sci. 81(4): 717-733, 2015.
- Z. Xie, G. Huang, J. He, Y. Zhang. A Clique-based WBAN Scheduling for Mobile Wireless Body Area Networks. In Journal of Procedia Computer Science, Vol. 31: 1092–1101, 2014.
- Z. Xie, G. Huang, R. Zarei, J. He, Y. Zhang. Wireless Sensor Networks for Heritage Deformation Detection and Tracking. Sensors, Vol. 14: 20562-20588, 2014.
- J. He, Y. Zhang, G. Huang, P.de Souza. CIRCE: Correcting Imprecise Readings and Compressing Excrescent points for querying common patterns in uncertain sensor streams. Information System. Volume 38(8): 1234-1251, 2013.
- X. Tao, Y. Miao, Y. Zhang, Z. Shen. Collaborative medical diagnosis through Fuzzy Petri Net based agent argumentation. Proceedings of 2014 IEEE International Conference on Fuzzy Systems, pp: 1197-1204. 2014.

Project ID

DP140100841

First Named Investigator

Yanchun Zhang

Scheme

Discovery Projects

Statement

Uploaded PDF file follows on next page.

Project ID: DP140100841, Deep Data Mining for Anomaly Prediction from Sensor Data Streams

CIs: Prof. Y. Zhang (Victoria University), Dr. Guangyan Huang (Deakin University)

The project was formally started in January of 2014. The agreements were finalized and fund received.

This project is to discover anomaly patterns from historical sensor data streams, based on which to predict anomalies from input sensor data streams. A research fellow, Dr Zhou has been recruited, who has a PhD in computer science and with strong research experience in the area of stream data mining, event detection over streams, media data management and mining.

The research progress achieved over the period is outlined below:

- We developed a new method of Finding Topic Clusters of Co-occurring Terms (FTCCT), which can automatically summarise documents by steady topics from a stream of short text. The quality topics that are loyal to the original meaning of text data ensure we can discover hot events and track their changes over time. FTCCT is a general method for sequences and can be easily transferred to process sensor data streams after changing sensor steams into sequences.
- We also provided a *generic Optimized Fuzzy Association Rule Mining* (OFARM) method for quantitative data, which are automatically generated continuous data (ie, a sequence of floating point numbers), such as medical sensor data and financial stream data. OFARM mainly resolves *sharp boundaries problem* in continuous data; that is, more precisely segment continuous data into meaningful/reasonable segments. Three datasets, including "Wisconsin Diagnostic Breast Cancer (WDBC)", "Wisconsin Prognostic Breast Cancer (WPBC)", "Pima Indians Diabetes" from UCI (University of California at Irvine) repository, have been used to demonstrate the effectiveness and efficiency of our OFARM algorithm.
- We developed an efficient agglomerative hierarchical clustering method for movie data, which exploits the advantages of both partitioned clustering and hierarchical clustering algorithms.

Publications

R. Zarei, J. He, G. Huang, Y. Zhang. Effective and efficient detection of premature ventricular contractions based on variation of principal directions. Digital Signal Processing, Vol. 50: 93-102, 2016.

S. Siuly, X. Yin, S. Hadjiloucas, Y. Zhang. Classification of THz pulse signals using two-dimensional cross-correlation feature extraction and non-linear classifiers. Computer Methods and Programs in Biomedicine, Vol. 127: 64-82, 2016.

X. Zhou, L. Cao, L. Chen, Y Zhang, G. Huang. Online Video Recommendation in Social Community, Proc. Of ACM SIGMOD Conference, pp: 1645-1656, Australia, 2015.

G. Huang, J. He, Y. Zhang, W. Zhou, H. Liu, P. Zhang, Z. Ding, Y. You, J. Cao. Mining Streams of Short Text for Analysis of World-Wide Event Evolutions. World Wide Web journal, Vol. 18 (5): 1201–1217, 2014.

H. Zheng, J. He, G. Huang and Y. Zhang. Optimized Fuzzy Association Rule Mining for Quantitative Data. Proceeding of IEEE International Conference on Fuzzy Systems, pp. 396-403, China, 2014.

Y. You, G. Huang, J. Cao, E.g Chen, J.g He, Y. Zhang, L. Hu. GEAM: A General and Event-Related Aspects Model for Twitter Event Detection. Proceeding of 14th International Conference of Web Information Systems Engineering, 319-332, China, 2013.

Project ID

LP100100624

First Named Investigator

Yuan Miao

Scheme

Linkage Projects

Statement

Uploaded PDF file follows on next page.

LP100100624, Data exchange and Service Integration with Applications in Heath Information systems

CIs: Yuan Miao, Yanchun Zhang

Data exchange and service integration of patient medical information is of clear benefit. However, such an application faces significant challenges including technical difficulties, financial affordability, and management concerns. This project aims to research a novel approach which includes a heterogeneous data exchange model, self aware schema mapping, web services coordinated localised autonomous services, cognitive models of medical information systems and human computer interaction. The new model of data integration and exchange will be developed with, and trialed by, Westgate General Practice Network, a large medical health provider.

The research has been completed and the team is preparing the final report.

- This project is to develop a data exchange and service integration model for heterogeneous data exchange and integration. We have developed an electronic referral and medical service providers' online communication system. This system has been high graded by partner organisation, Westgate General Practice Network.
- For evidence based resource allocation, we have designed a one key data collection model, and will implement a prototype in the coming six months, intenerate it with the most widely used GP electronic medical record system: Medical Director. The internal test has shown that the model is feasible and practical.
- For knowledge based service integration, we discovered that there is a need for medical service providers to model their knowledge easily without mediation of computing specialists. Yet existing computing models could not meet this requirement. We have developed a visualised fuzzy cognitive map model for visualising knowledge by domain experts. Initial investigation on modeling several existing published knowledge on medical decision support has shown its power and improvement in visualising knowledge by domain experts.

Publications

1. X.H. Tao, Y. Miao, Y. C. Zhang, Z. Q. Shen. Collaborative Diagnosis through Fuzzy Petri Net Based Agent Argumentation. Proceeding of 2014 IEEE International Conference on Fuzzy Systems, pp: 1197-1204, 2014.

2. Y. Miao. Modelling Dynamic Causal Relationship in Fuzzy Cognitive Maps. Proceeding of 2014 IEEE International Conference on Fuzzy Systems, pp: 1013-1020, 2014.

3. Y. Miao. Fuzzy cognitive map for domain experts with no artificial intelligence expertise. Proceeding of the 8th International Conference on Control, Automation, Robotics and Vision, pp: 486-492, 2014.

4. D. Toro, H. Xu, Y.H. and Y. Miao. Health Benefits of Wheelchair Tai Chi and Spinal Cord Injuries. Proceeding of Australasian Acupuncture & Chinese Medicine Annual Conference, pp: 15-27, Australia, 2014.

5. Y. Miao. Visualising Fuzzy Cognitive Maps. Proceeding of 2012 IEEE International Conference on Fuzzy Systems, pp: 1-8, 2012.

6. X.H. Tao, Y. Miao, Y. C. Zhang. Cooperative-competitive healthcare service negotiation. International Journal of Software and Informatics, vol. 6, no. 4, pp: 553-570, 2012.

7. Y. Miao, C. Y. Miao, X.H. Tao, Z.Q. Shen, Z.Q. Liu. Transformation of Cognitive Maps. IEEE Transactions on Fuzzy Systems, Vol 18, No. 1, pp: 114-124, 2010.

8. X. B. Yang, Y. Miao and Y.C. Zhang. Model-driven GUI automation for efficient information exchange between heterogeneous electronic medical record systems. Information Systems Development: Business Systems and Services: Modeling and Development, Springer New York, pp: 799 – 810, 2011.

Project ID

LP100200682

First Named Investigator

Yanchun Zhang

Scheme

Linkage Projects

Statement

Uploaded PDF file follows on next page.

Project ID: LP100200682, Real-time and Self- Adaptive Stream Data Analyzer for Intensive Care Management.

CIs: Y. Zhang (VU), X. Yi (VU), J. He (VU), M. Steyn (RBWH), K. Taraporewalla (RBWH) and J. Cao (NUFE)

This research project focused on both fundamental and application-oriented research. The fundamental research includes: 1) time series data compression and correction; 2) feature extraction; and 3) modelling of generative process of physiological time series. Specifically, we developed: 1) algorithms that can greatly reduce the data size while retaining critical information for data analysis; 2) several feature extraction algorithms which can better capture the intrinsic characteristics of the original data, and build more compact and representative dataset; 3) models of generative process of physiological time series, which foster a greater understanding of the underlying intrinsic dynamics of physiological time series. The application-oriented research solves the challenges encountered by the medical industry in task-specific applications. Specifically, this project developed models and algorithms for anomaly detection, future event prediction, time series classification, and time series clustering. By utilizing these research works, medical specialists can reduce the response time and better diagnose the patients. Hence the workload of the clinicians' can be reduced, and their medical performance can be enhanced in terms of higher success rate and lower mortality rate in surgery and intensive care unit.

We also investigated the practical issues that impede adoption of data mining techniques in the healthcare industry, and proposed a new healthcare data consumption model, a cloud-based healthcare data mining service framework, and a multi-agent computing model for online collaborative medical diagnosis. These works lay foundation for future development of healthcare data mining services and promote adoption of data mining techniques in the medical industry. The techniques developed in this project have broad areas of potential application. They will be readily embraced in other domains as diverse as financial systems, network monitoring, security, telecommunications data management, web applications, manufacturing, and sensor networks.

The project is completed and the final report is going to submit to the ARC. This project produced 24 research papers (15 journal papers & 9 conference papers, detailed in the final report), and trained two PhD students. The research has received high profile media coverage including in The Age (print circulation of 157,486 + online) on 11 March 2013 and The Australian IT section (print circulation of over 122,428 + online) on 26 March 2013. A Chinese version of the article from The Age was re-published in over 15 Chinese media channels on the same week, including in the popular ChinaDaily, XinhuaNet and ChinaNews.

We select a few high quality published papers supported by the project:

- G. Huang, Y. Zhang, J. Cao, M. Steyn, K. Taraporewalla, Online mining abnormal period patterns from multiple medical sensor data streams. World Wide Web 17(4): 569-587 (2014) [ERA ranking A]
- Siuly, E. Kabir, H. Wang and Y. Zhang, Exploring sampling in the detection of multi-category EEG signals', Computational and Mathematical Methods in Medicine, 2015, Article ID 576437, 12 pages, http://dx.doi.org/10.1155/2015/576437.[*ERA Ranking A**]
- Siuly and Y. Li, Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification, Computer Methods and programs in Biomedicine, Computer Methods and programs in Biomedicine, 2015 (119): 29-42. [*ERA Ranking A**]
- Supriya, S., Siuly, S., Wang, H., Cao, J., & Zhang, Y. Weighted Visibility Graph With Complex Network Features in the Detection of Epilepsy. IEEE Access, 2016 (4): 6554-6566. [*ERA Ranking A**]
- Yin, X. X., Hadjiloucas, S., Zhang, Y., Su, M. Y., Miao, Y., & Abbott, D. Pattern identification of biomedical images with time series: contrasting THz pulse imaging with DCE-MRIs. Artificial intelligence in medicine, 2016(67): 1-23. [*ERA Ranking A**]

Project ID

LP150100673

First Named Investigator

Yanchun Zhang

Scheme

Linkage Projects

Statement

Uploaded PDF file follows on next page.

Project ID: LP150100673, Privacy Preserving Data Sharing in Electronic Health Environment.

CIs: Y. Zhang (VU), H. Wang (VU), X. Yi (RMIT), J. Soar (USQ), Y. Miao (VU), E. Bertino (Purdue U) and B. Heyward (Nexus on-line).

The agreement between the ARC and all partners was signed in July 2016. The team has started to work for the project from 2015 and a research assistant has already involved in the project in 2016. We are working on the development of an EHD system that manages a federation of EHD databases from a large number of medical organisations. Access to any EHD needs joint authentication from t out of n parties. We plan the following approach to achieving n server joint authentication over one database: (1) Given G as a group of a large prime order q with a generator g, each server randomly chooses a private key sk_i from Z_q, and computes the corresponding public key $pk_i=g^{ski}$; then we can produce the joint authentication PK key as the product of the individual public keys. (2) Encrypted EHDs: Suppose each EHD is represented as m; we store $A=g^r$, and $B=mPK^r$, which are encrypted health records. (3) Even if the EHD is protected with n server joint authentication, we do not want them to collude and allow flexible access because users may be doctors from many organisations.

The following high quality papers have been published.

S. Supriya, S. Siuly, H. Wang, J. Cao and Y. Zhang. Weighted Visibility Graph With Complex Network Features in the Detection of Epilepsy. IEEE Access, vol. 4, pp: 6554-6566, 2016. DOI: 10.1109/ACCESS.2016.2612242

P. Vimalachandran, H. Wang, Y. Zhang, G. Zhuo. The Australian PCEHR System: Ensuring Privacy and Security through an Improved Access Control Mechanism. EAI Endorsed Transactions on Scalable Information Systems 16(8): e4. 2016. DOI: 10.4108/eai.9-8-2016.151633

S. Badsha, X. Yi, I. Khalil. A Practical Privacy-Preserving Recommender System. Data Science and Engineering 1(3): 161-177, 2016.

S. Subramani, H. Wang, S. Balasubramaniam etc. Mining Actionable Knowledge Using Reordering Based Diversified Actionable Decision Trees. Web Information Systems Engineering, LNCS, Vol. 10041, pp: 553-560, 2016.

M. Peng, J. Zhu, X. Li, J. Huang, H. Wang, and Y. Zhang. Central Topic Model for Event-oriented Topics Mining in Microblog Stream. CIKM '15, pp: 1611-1620, Australia, 2015.

J. Huang, M. Peng, and H. Wang. Topic Detection from Large Scale of Microblog Stream with High Utility Pattern Clustering. PIKM '15, pp: 3-10, Australia, 2015.