Security-oriented Data Grids for Microarray Expression Profiles

Richard SINNOTT [a,1] Christopher BAYLISS [a] Jipu JIANG [a]
[a] *National e-Science Centre,University of Glasgow,United Kingdom*

Abstract. Microarray experiments are one of the key ways in which gene activity can be identified and measured thereby shedding light and understanding for example on biological processes. The BBSRC funded Grid enabled Microarray Expression Profile Search (GEMEPS) project has developed an infrastructure which allows post-genomic life science researchers to ask and answer the following questions: who has undertaken microarray experiments that are in some way similar or relevant to mine; and how similar were these relevant experiments? Given that microarray experiments are expensive to undertake and may possess crucial information for future exploitation (both academically and commercially), scientists are wary of allowing unrestricted access to their data by the wider community until fully exploited locally. A key requirement is thus to have fine grained security that is easy to establish and simple (or ideally transparent) to use across inter-institutional virtual organisations. In this paper we present an enhanced security-oriented data Grid infrastructure that supports the definition of these kinds of queries and the analysis and comparison of microarray experiment results.

Keywords. Microarray Expression Profiles, Grid security, Shibboleth

Introduction

The UK Biotechnology and Biological Research Council (BBSRC) Grid Enabled Microarray Expression Profile Search (GEMEPS) project (www.nesc.ac.uk/hub/projects/gemeps) began in March 2006. The fundamental premise upon which GEMEPS is based is that life scientists recognise that it is to their advantage to collaborate, especially with regard to sharing of expensively produced microarray experiments. Academics and researchers will always need to refer to and publish in journals and leading publications in their respective fields, however targeted real time access to research data between collaborators and institutes needs to occur to expedite the knowledge discovery process. Currently this is largely not the case and access to and usage of microarray data sets is limited for a variety of reasons: competitive, ethical, social, political being just a few. To support any form of data sharing models, scientists and their supporting IT staff need technologies that allow them to be fully informed and in control of the security infrastructures by which they make their data sets available and to whom.

The Grid in principle provides an appealing model for access to and usage of distributed and heterogeneous life science data sets. The explosion of data sets across the

[1] r.sinnott@nesc.gla.ac.uk

life science spectrum, and the major compute demands of high through post-genomic research offer direct requirements siuitable for Grid based solutions. However Grid technologies are not a silver bullet or a complete panacea for all of the challenges facing the life science community. The Grid needs agreements and standards on how life science data sets are created, defined and annotated before it can be exploited for data discovery, analysis or linkage. Similarly, understanding of the life science applications and data sets and the specific requirements they impose on computational resources is needed before the Grid can truly solve the compute requirements of this community.

Perhaps the most important aspect to recognise is that technology alone is insufficient to solve the requirements from this domain and must be guided by the wider scientific community needs and experiences. It could be argued that there has primarily been a middleware push as opposed to a scientific pull across the majority of the Grid research communities, and this is especially so in the life sciences. Previous projects such as the DTI funded Biomedical Research Informatics Delivered by Grid Enabled Services (BRIDGES) (www.nesc.ac.uk/hub/projects/bridges) at the National e-Science Centre (NeSC) at the University of Glasgow and funded reports such as [1] have identified that life scientists are especially wary of their data resources being accessed by others without them first exploiting their results, e.g. through journal publication. This cultural issue is especially significant since technologies must be met by scientific willingness to engage and collaborate. Yet the existing Grid security solutions are largely complex and confusing to end users and the supporting IT staff. Thus technologies are needed which simplify as much as possible the access to and usage of a range of data sets and resources more generally. A key and crucial benefit of the Grid is to support site autonomy. Sites should be able to define and enforce their own local policies on access to and usage of data sets. Since large scale post-genomic scientific research is rarely undertaken by a single site, but requires access to a range of data sets and resources including public repositories as well as collaborators private resources, multi-site solutions are needed. The definition of these security policies also needs to be recognised across the multiple institutions involved in collaborative research.

Importantly the scientific community needs to be made aware of what it means to provide controlled access to their research data and the potential ramifications thereof. Biologists tend not to be computer scientists and are unfamiliar with advanced Grid data access or security solutions. As such any solutions that are put forward in this domain have to be intuitive and allay their potential fears on compromises of their research data, or potential exploitation by competitors or third parties. New developments such as gene identification, gene function and development of new targeted drugs offers enormous opportunities for researchers both academically and commerically. As such, they need to be completely satisfied that any new technological solutions will fit into the way in which they wish to work, and importantly protect their research results and data from compromise.

In this paper we outline the solutions developed at the National e-Science Centre (NeSC) at the University of Glasgow to support seamless Grid based access to a range of services that allow discovery, analysis and comparison of microarray experiments.

The rest of the paper is structured as follows. Section 1 outlines the background to microarray data sets and associated standards. Section 2 focuses on the architecture and implementation of the GEMEPS infrastructure focusing in particular on security aspects

and services to ascertain microarray experiment similarity. Finally in section 3 we present our conclusions and plans for the future.

1. Background to GEMEPS

The GEMEPS project aims to develop a Grid infrastructure for discovery, access, integration and analysis of microarray data sets. Through the GEMEPS infrastructure scientists should be able to ask the following kinds of questions and obtain appropriate results based upon their privilege:

- who has run a microarray experiment and generated similar results to mine?
- who has undertaken experiments and produced data relevant to my own interests, e.g. for a particular phenotype, for a particular cell type, for a particular pathogen, on a particular platform or microarray chip set?
- show me the results from a particular collaborator;
- show me the conditions and analysis associated with experimental results similar to mine.

In all of these scenarios, the model we consider is for sites to keep and maintain their own data and define their own security policies on access and usage. This model has a psychological benefit to encourage collaboration, namely that scientists are not simply making their data publicly available for example in one of the existing repositories such as Gene Expression Omnibus (GEO) at NCBI [2], ArrayExpress [3] or CIBEX [4]. Scientists are often reluctant to publish their data in such repositories until they have published results in recognised journals which can, depending on the journal be a long and protracted affair. As a result, these public repositories tend to be populated with older data sets. It is also the case that these data repositiories provide various kinds of services through which the repositories themselves might be searched or mined. These repositories typically require data sets to be MIAME compliant [5]. The stated goal of MIAME is to outline the minimum information required to interpret unambiguously and potentially reproduce and verify an array based gene expression monitoring experiment. Whilst the details of particular experiments may be different, it is the intention of MIAME to define a core that is common to most experiments. MIAME is not a formal specification, but a set of guidelines concentrating on the content of information and various metadata that needs to be captured to facilitate re-use or reproduction of experimental results. Most major journal publications now require data associated with journal papers to be published in combination with the paper itself.

A MIAME description typically describes the design of: array platform - containing the description of the common features of the array and the description of each array design element; gene expression experiment - containing a description of the overall experimental design; the samples used; how extracts were prepared; which hybridisation procedures were followed and ultimately what data was measured and how it was analysed and normalised.

MIAME compliance is not prescriptive in the sense that all or a given subset of the various sections that might be associated with a given experiment must be given. These sections are usually provided in free text format, along with recommendations requiring maximum use of controlled vocabularies or external ontologies. MIAME recognises that

few controlled vocabularies have been fully developed, hence it encourages users to provide their own qualifiers and values identifying the source of the terminology. Of those that are available, the Microarray Gene Expression Data Society (MGED) [6] is one of the more established ontologies for microarray experiment description.

Several data formats have also been defined and applied across different sites with different user communities. These include: Microarray Gene Expression Markup Language (MAGE-ML) [7] is part of the MGED family of standards and is MIAME compliant and XML based. Many major repositories, such as GEO, ArrayExpress and CIBEX support results being deposited in MAGE-ML as well as supplying data in that format. Simple Omnibus Format in Text (SOFTtext) [8] is a simple text based format designed by GEO. Unlike MAGE-ML, SOFTtext is not XML based using instead keywords for describing platform, sample and results. It has fewer fields than MAGE-ML yet is still MIAME compliant. GEO supports submissions in this format and makes results availiable in it as well. Since SOFTtext is based around a simple format it is easy to parse and use. MIAME Notation in Markup Language (MINiML) [9] is an XML based format used by GEO and is equivalent to SOFTtext. The NCBI accepts data deposited in MINiML format and makes records available in this format. MINiML can be considered an XML equivalent to SOFTtext as it provides the same properties, however in XML form. NCBI has made a schema for MINiML avaliable allowing a validating parser to confirm that a MINiML file is well formed. This is a distict advantage over SOFTtext where there is no formal definition of how the files should be formatted. As with the other SOFT formats MINiML is MIAME compliant yet has fewer fields than MAGE-ML. The relative simplicity of MINiML when compared to MAGE-ML has direct advantages for usability and associated learning curve.

SOFTmatrix [10] is a new format developed by NCBI based on MIAME and using Microsoft Excel spreadsheet templates. Excel spreadsheets are one of the most common ways in which scientists keep their own microarray experimental results.

As seen a multitude of on-going efforts in how to describe and annotate the data and metadata associated with microarray experiments and results exist. It is within this context that the GEMEPS project is developing a security oriented Grid infrastructure for discovery and comparison of microarray experiment profiles.

2. Secure Data Grids within GEMEPS

The Open Middleware Infrastructure Institute (OMII - www.omii.ac.uk) and the Open Grid Service Architecture - Data Access and Integration (OGSA-DAI - www.ogsadai.org.uk) technologies were applied to produced an alpha prototype showing how Grid data access middleware could be used for access to and usage of microarray data sets (shown in the striped boxes). This was primarily used for feasibility studies.As stated, GEMEPS recognises and distinguishes between public microarray resources and private resources such as those created and maintained at the Sir Henry Wellcome Functional Genomics Facility (SHWFGF) at the University of Glasgow. The SHWFGF has been the primary source of requirements driving GEMEPS development.

2.1. Security Aspects of GEMEPS

To simplify the end user experience in accessing and using this resource and in providing advanced Grid security, we have provided Shibboleth based single sign-on [11,12] The Shibboleth technologies are developed by the Internet2 community and offer a simple yet secure way in which single sign on to a variety of distributed resources can be supported. The Shibboleth framework provides a mechanism for exchanging attributes across different organisations for the purpose of authentication and authorisation. It enables a user to access a protected resource or service at a remote domain (commonly referred to as a Service Provider (SP) or target) by using the user's own home security domain (commonly referred to as an Identity Provider (IdP) or origin) to perform user authentication. The framework uses X.509 certificates for the underlying secure attribute exchange. An important advantage the framework provides is that the user is not required to possess an X.509 certificate. This is because Shibboleth allows inter-institutional sharing of resources within a trusted federation where it is the responsibility of the home institutions to authenticate their users. Therefore Shibboleth directs the users to their home institution for authentication. The information which is exchanged as attributes helps to determine whether to grant the user access to the resource at the SP. To achieve this Shibboleth uses Security Assertion Mark-up Language (SAML) [13,14], an OASIS standard for exchanging authentication and authorisation statements, between the IdP and the SP. When the user is authenticated, the Shibboleth component at the SP establishes and maintains a session with the user's web browser on behalf of the resource the user is accessing. This session consists of cookies which are passed between the web browser and web server. The cookies are associated with a security context which holds the user's authentication information and a set of attributes describing the user's identity. With this, the user can access the resource (or resources across the federation/virtual organisation) more than once without repeating the Shibboleth authentication process until the cookies expire or are deleted from the user's machine.

The Shibboleth framework enables the creation of a federation to build trust relationships between participating organisations for inter-institutional access of resources. These organisations exchange attributes using the Shibboleth protocols and abide by a common set of policies and practices. Currently within GEMEPS Shibboleth access to the portal is based upon a single IdP (at the University of Glasgow) however extensions to this to allow access by other users from remote institutions are feasible and work is on-going in this area.

The Shibboleth framework devolves responsibility of user authentication to the user's home institution. This avoids the need to create a separate authentication system that is exclusive to the GEMEPS system, e.g. a portal log-in. It is of course possible to set up single usernames and passwords for the GEMEPS portal however this is not an especially scalable solution since users collect many usernames and passwords in the course of their research and often keep the same username/password combination to minimize the amount of information they have to memorize. Furthermore, the Shibboleth framework also provides a scalable and an extensible solution for managing access to resources. By using the Shibboleth framework it is possible to accommodate a growing number of users from different institutions as part of a federated access management arrangement. The Shibboleth framework is also being adopted by a number of higher educational institutes in the UK to develop the next generation access management system

for their users. Indeed the UK federation was established in November 2006, and many other federations are being established internationally.

Once a user has successfully authenticated at their home institution and in turn through Shibboleth they are authenticated to the GEMEPS portal they are presented with a variety of services (portlets). We note that which attributes are released from a given IdP to a particular SP is configurable and dependent upon the attribute release and attribute acceptance policies. Through the NeSC DyVOSE project (www.nesc.ac.uk/hub/projects/dyvose) we have implemented solutions which exploit delegation of authority to allow the dynamic creation or revocation of these attributes across multiple sites [15,16]. Alternatively the roles themselves and how they are assigned to users at different sites can be achieved in a variety of other ways. For example, if their authentication system is based upon LDAP then an LDAP editor can be used to simply add these roles to the particular users by the local system administrator. UK academia more generally has identified and recommended four key attributes based around the eduPerson object class [17]. These will be used to determine that an individual is a member of a recognized UK HE/FE institution. Often this is sufficient information for a service provider to allow/deny access, e.g. e-Journals might only need to know that a user is a member of a given university to allow/deny access (based on whether that institution has a subscription to that particular journal say).

The returned attributes are used to configure and customize (authorize) the resources that are accessible to the user. Thus here the attribute *gemeps_human_genome_researcher* is here used to allow access to human genome related experiments. Typically the management and organization of these roles and their relationship to security policy is achieved through some form of coordinated access control policy. Role Based Access Control (RBAC) technologies such as PERMIS (www.permis.org) can be used to enforce authorization policies and manage hierarchies of roles and their associated privileges. Practical explorations of these technologies are described in detail in [18,19,20].

We note that several portlets are available within this portal for querying the status of the Grid infrastructure, for running large scale bioinformatics BLAST applications on a variety of large scale HPC facilities. The GEMEPS portlet allows for discovery and comparison/similarity checking of microarray experiments. This is achieved firstly through meta-data querying, and then for rank correlation evaluation.

2.2. Meta-data Query Services

The various mark-up languages associated with microarray experiments allow for capturing and annotation of a variety of information about particular microarray experiments. Examples of the kinds of information that are typically captured include the specification of: how the data was processed; the type of molecules extracted from the biological material; the unique ID of the samples; descriptions of the scanning and image acquisition protocols (both hardware and software); the conditions used to grow or maintain organisms or cells prior to extract preparation; the name of the company, laboratory or person that provided the biological material; the protocols used to isolate the extract material; the platform used, e.g. GPL570 for the human genome; the species under study, e.g. homo sapiens, rattus rattus, etc; the biological material and the experimental variable(s) for the sample; a description of the experiment more generally and the protocols used for hybridization, blocking and washing, and any post-processing steps used such as staining.

Scientists in the first instance would like to be able to query across a range of experiments based on any one or more of these kinds of search terms. Through such queries immediate and meaningful experimental results can be returned. Thus scientists are unlikely to be interested comparing experimental results from homo sapiens and barley for example. We note that at the gene name level however, it is often the case that common gene name clashes do exist across species for example. To support this basic metadata querying, the GEMEPS project has implemented a simple user oriented portlet that allows for a variety of these kinds of information to be used for querying over available (subject to authorisation privileges) data sets.

2.3. Similarity Services

Having identified the set of experiments that are relevant to a researcher, finer grained analysis and comparison is needed to understand how relevant these data sets actually are. To support this GEMEPS has explored two different rank correlation cooefficient algorithms based upon Spearman Rank [21] and Kendall Tau [22]. The implementation of these algorithms is facilitated by the building of indexes from microarray experiments. Amongst the challenges associated with these algorithms in this domain are the difficulties in identifying and relating gene names across different experiments. Different experiments might use their own naming schemes, different ontologies such as MGED. Within the course of the project we have explored Life Science Identifiers (LSids) to address these issues. LSids are designed as Uniform Resource Names (URN) written in the form: urn:lsid:<authority>:<database>:<object>:<version> where <authority> is the name of the authority who issued the LSid, <database> is the name of the authority's database the LSid is stored in and <object>:<version> identifies the object within the database and its revision.

LSids are intended to serve as persistent identifiers allowing them to be used without later being reassigned. They allow to map to exactly the same set of bytes permanently. This means that an LSid, once assigned, is permanently attached to a specific encoding of its data which cannot be updated or corrected. An immediate advantage of this is that makes LSids usable as references. The LSid specification suggests using an LSid proxy, e.g. lsid.biopathways.org, to resolve LSids. The biopathways resolver provides LSids for many existing data sets such as the NCBI databases, ArrayExpress and SwissProt for example.

At the time of writing, it is unclear whether LSids will solve the problems arising in uniquely identifying information in the life science domain. For example, the closure of the Interoperable Informatics Infrastructure Consortium (i3c) means the loss of RDF metadata associated with LSids. References to this data still appear in examples and tutorials but the i3c itself website no longer exists. The only implementations of the LSid stack found are from the IBM LSid project on sourceforge. The logs of the source repository reveal little activity with the majority of the code remaining untouched since 2004.

To address this, the project has focused upon developing solutions targeted towards SOFTtext and MINiML. The implementation of these algorithms produces initial results as would be expected. Thus for example, when the results from one experiment are compared with itself a correlation co-efficient of 1 is returned. When the inverse of the results of an experiment, i.e. reversing the gene expression ranking, the algorithm returns

-1 as expected. The implementation itself offers currently just the Spearman ranking co-efficient however the Kendall Tau correlation co-efficient is also under development and will be rolled out in due course.

The interface to this system allows users to either upload their own experimental results for comparison from a file (given as a sequence of gene names, or as a sequence of <gene name, expression value> orderings), or they can cut and paste these, e.g. from an Excel spreadsheet - a common technology used by most life science researchers in managing their microarray results. The final result of these experiments are given as a ranking of most similar experiments. The most similar given with the level of similarity (Spearman rank coefficient) given. With the most similar experiment identified, the end user may then follow the hyperlink to obtain more information about this particular data and how the results were obtained etc.

3. Conclusions

At the time of writing the GEMEPS project has been on-going for 10 months and has a further 2 months remaining. The project has faced many challenges in reaching the current implementation status. Perhaps the greatest of these is in understanding, managing and linking the bioinformatics data resources. The lack of a common naming system and a variety of different mark-up languages and ontologies to describe the results of microarray experiments adds to the overall complexity in development of Grid based systems.

We believe that the adoption and roll out of Shibboleth to simplify the access to and usage of secure bioinformatics data resources is key to the success of this and other solutions. Allowing scientists to remain autonomous and keep their own datasets locally but allow secure and controlled access by remote collaborators offers a much more appealing model to scientific data sharing than centralised public microarray repositories. Trust plays an important role in Shibboleth (or any security based system). Sites trust remote institutions to authenticate their users correctly. To address this the University of Glasgow has rolled out a unified account management system based on Novell nSure technology. With this system, there is a one-one mapping between members of the university - either staff or students - and the privileges that they possess. Thus users do not have different usernames and passwords for different systems. Through this system, when a member of staff or student leaves, all of their privileges are removed. This system is being rolled out as part of the JISC funded Glasgow early adoption of Shibboleth (GLASS) project (www.nesc.ac.uk/hub/projects/glass).

The work on GEMEPS is still on-going and we are now focusing on several areas. These include hardening and scaling the existing system. Thus for example, the Spearman Rank algorithm does not scale well when large sets of genes in one experiment are not available (being expressed) in another. This might at first instance imply that the two experiments are disimilar hence should have a low correlation co-efficient. However it is often the case that biological meaning from microarray experiments should only be determined from the most significantly expressed genes. Thus below a given limit, the expression values and associated gene orderings cannot be guaranteed due to statistical variation when conducting the experiments. To address this we are exploring various cut off scenarios for experiment comparison. Thus a scientists might only be interested

in the top 10, 100, 1000 expressed genes. Alternatively a scientist might only be interested in genes expressed where the expression value itself is above a given cut off. These combinations and their impact of result accuracy offer important ways to increase biological understanding of the accuracy of microarray experiments. A further scenario we are exploring is where a scientist is interested in experiments where a particular gene is expressed or expressed above a particular varlue.

A further enhancement to this system is to support large scale expression profile matching. Thus when many thousands of matching experiments are returned from metadata queries and need to be compared with one or more experiments, then use of large scale HPC facilities is needed. To address this we are exploring user driven access to and usage of large scale HPC facilities such as the UK e-Science National Grid Service (www.ngs.ac.uk) and the ScotGrid system (www.scotgrid.ac.uk) at Glasgow. In achieving this, the Grid will be completely shielded from the user.

4. Acknowledgements

The work described here was suported by a grants from the UK Biotechnology and Biological Sciences Research Council (BBSRC). We gratefully acknowledge their support. We also acknowledge inputs to the science we are trying to support from Dr Pawel Herzyk at the Sir Henry Wellcome Functional Genomics Facility at the University of Glasgow. Dr John Watt is acknowledged for his Shibboleth expertise and support.

References

[1] P. Lord, A. MacDonald, R.O. Sinnott, Large-scale data sharing in the life sciences: Data standards, incentives, barriers and funding models: The "Joint Data Standards Study", prepared for The Biotechnology and Biological Sciences Research Council, The Department of Trade and Industry, The Joint Information Systems Committee for Support for Research, The Medical Research Council, The Natural Environment Research Council and The Wellcome Trust.

[2] Gene Expression Omnibus (GEO), www.ncbi.nlm.nih.gov/geo/

[3] P. Rocca-Serra, A. Brazma, H. Parkinson, U. Sarkans, M. Shojatalab, S. Contrino, J. Vilo, N. Abeygunawardena, G. Mukherjee, E. Holloway, M. Kapushesky, P. Kemmeren, G. Garcia Lara, A. Oezcimen, S.-Assunta Sansone. ArrayExpress: a public database of gene expression data at EBI. C R Biol, 326(10-11):1075–1078, Oct 2003.

[4] K. Ikeo, J. Ishi-i, T. Tamura, T. Gojobori, and Y. Tateno. CIBEX: center for information biology gene expression database. C R Biol, 326(10-11):1079–1082, Oct 2003.

[5] Minimal Information About a Microarray Experiment (MIAME), http://www.mged.org/Workgroups/MIAME

[6] Microarray Gene Expression Data Society (MGED) Ontology Working Group, http://www.mged.org/ontology

[7] MicroArray and Gene Expression Markup Language (MAGE-ML), http://www.mged.org/Workgroups/MAGE/mage-ml.html

[8] Simple Omnibus Format in Text (SOFTtext), http://www.ncbi.nlm.nih.gov/projects/geo/info/soft2.html

[9] MIAME Notation in Markup Language (MINiML), http://www.ncbi.nlm.nih.gov/projects/geo/info/MINiML.html

[10] Simple Omnibus Format in Matrix (SOFTmatrix), http://www.ncbi.nlm.nih.gov/projects/geo/info/soft2.html

[11] Shibboleth Architecture Technical Overview, http://shibboleth.internet2.edu/docs/draft-mace-shibboleth-tech-overview-latest.pdf

[12] Shibboleth Architecture Protocols and Profiles, http://shibboleth.internet2.edu/docs/draft-mace-shibboleth-arch-protocols-latest.pdf

[13] OASIS, Assertions and Protocol for the OASIS Security Assertion Markup Language (SAML) v1.1, 2 September 2003, http://www.oasis-open.org/committees/security/

[14] Security Assertion Markup Language (SAML) version 2.0, March 2005, http://www.oasis-open.org/specs/index.php#samlv2.0

[15] R.O. Sinnott, J. Watt, O. Ajayi, J. Jiang, J. Koetsier, A Shibboleth-Protected Privilege Management Infrastructure for e-Science Education, 6th IEEE International Symposium on Cluster Computing and the Grid, CCGrid2006, May 2006, Singapore.

[16] R.O. Sinnott, J. Watt, O. Ajayi, J. Jiang, Shibboleth-based Access to and Usage of Grid Resources, IEEE International Conference on Grid Computing, Barcelona, Spain, September 2006.

[17] eduPerson Specification, http://www.educause.edu/eduperson/

[18] R.O. Sinnott, Security Focused Federation of Distributed Biomedical Data, Proceedings of UK e-Science All Hands Meeting, 4-6 September 2003, Nottingham, England.

[19] R. O. Sinnott, M. M. Bayer, J. Koetsier, A. J. Stell, Advanced Security on Grid-Enabled Biomedical Services, Proceedings of UK e-Science All Hands Meeting, September 2005, Nottingham, England.

[20] R. O. Sinnott, M. M. Bayer, J. Koetsier, A. J. Stell, Grid Infrastructures for Secure Access to and Use of Bioinformatics Data: Experiences from the BRIDGES Project, 1st International Conference on Availability, Reliability and Security, (ARES'06), Vienna, Austria, April, 2006.

[21] Griffiths, D., A Pragmatic Approach to Spearman's Rank Correlation Coefficient, Teaching Statistics 2, pp. 10-13, 1980.

[22] Kendall, M. (1948) Rank Correlation Methods, Charles Griffin & Company Limited