Aquerre Technologies

(U) Automatic Graph Learning Algorithms

A PROPOSAL TO THE
NATIONAL SECURITY AGENCY
INNOVATIVE MISSION CAPABILITIES FY15 BAA
VOLUME I: TECHNICAL

BAA Number:	(U) BAA-IMC-15
Track:	(U) 1. Data Discovery, Analysis, Exploitation, and Management
Topics:	(U) 1. Behavioral Analytics 2. Contextual Discovery 3. Alternate Data Exploration Techniques 4. Predictive Analytics 5. Data Discovery Capabilities 7. Multi-source Data Fusion 10. Predictive Inferences
Proposal Title:	(U) Automatic Graph Learning Algorithms
Name of Offerer:	(U) Aquerre Technologies LLC
Address of Offerer:	(U) 1445 Colby Ave #3 Los Angeles, CA 90025-7844
Technical Point of Contact:	(U) Dr. Noah B. Jacobsen E-mail: noah@aquerre-technologies.com Telephone: 310-857-8049 Fax: 424-270-0709 Mail: 1445 Colby Ave #3 Los Angeles, CA 90025-7844
Contracts Point of Contact:	(U) Same as Technical PoC.

(U) Automatic Graph Learning Algorithms

Program: Innovative Mission Capabilities FY15
Track 1: Data Discovery, Analysis, Exploitation, and Management
Firm: Aquerre Technologies LLC

Contents

List of Figures

List of Tables

1 (U) Abstract

(U//FOUO) The Government has identified a strategic need for advanced data discovery, analysis, exploitation, and management technologies. In this response to the Innovative Mission Capabilities (IMC) Broad Agency Announcement (BAA) Fiscal Year (FY) 2015, we develop and deliver to the Government a Software Prototype of innovative, proprietary algorithms for predictive information processing for arbitrary numeric data systems and large data sets.

(U//FOUO) Predictive information science is widely applied across the technology markets, including financial, automotive, communications and business intelligence markets and advanced products are required to pace the growth in data and technology. The Prototype developed in this contract leads the state-of-the-art in predictive information systems by leveraging a powerful graphical framework for modeling the dynamical system of interest, coupled with innovative statistical learning and inference processing methods developed exclusively by Aquerre Technologies.

(U//FOUO) We will use historical daily stock data from Nasdaq.com plus historical daily climate data from NOAA.gov to demonstrate the capabilities of our Prototype of statistical learning and prediction algorithms. Our proprietary graphical modeling strategy and methods for estimating local probability distribution functions from training data enable the implementation of predictive information processing algorithms using variations of the well-known sum-product algorithm. We will demonstrate that the Prototype can estimate the state of hidden data (*e.g.* future data) based on automatically constructed graph-based statistical models from an arbitrary collection of dependent numeric training data. A major technical contribution of our software Prototype is the use of innovative proprietary graph learning algorithms based on the factor-graph modeling framework. The Prototype will feature automatic learning by adapting its system model automatically to training data, in addition to user-based model input. Our final demo and report will include comparison of our automatic graph learning methods to the Kalman filter (with manually defined parameters) as a performance benchmark.

2 (U) State of the Art

(U//FOUO) Quantitative prediction is a classical engineering problem with ubiquitous techniques such as linear regression [1, 2] and modern methods Bayesian estimation and prediction [3, 4], and neural networks and machine learning [5, 6]. For example, Bayesian predictive methods (using Bayes' rule to determine the posterior distribution of unobserved data given the observed data) are applied in radio-communication systems to estimate the transmitted data with near Shannon-capacity performance, in [4].

(U//FOUO) In this IMC R&D program we utilize a graphical modeling framework to develop an advanced statistical learning and prediction software Prototype. The graphical framework that we utilize (factor-graphs) is the same underlying framework that is used in the development of high-performance Error Correcting Codes (ECCs) (see *e.g.* [7]), such as the Low Density Parity Check (LDPC) code. The principal investigator is the inventor of a high-performance LDPC codec published in the IEEE Vehicular Technology Conference [8] and patented in the U.S. in 2011 [9]. A key innovation of the predictive software Prototype delivered to the Government in this IMC contract is the use of new and innovative methods for estimating the local probability functions (factors) using training-data based measurements of local statistical distributions (technical details are provided in Section 3.2.2). We further introduce reduced complexity methods based on sparse factor-graph representations for large systems.

(U//FOUO) Our concept system uses a plurality of input data-types for predictive information processing, however, in this three month contract we narrow the focus to a more tractable scope by considering only numeric time-series data from two disparate sources. With the narrowed focus to numeric data sources, we anticipate making rapid progress towards delivery of a basic operational Prototype in the three month work plan. We will include a performance comparison of our Prototype system to the well-known Kalman filter (*e.g.* [2]). In addition to the basic Prototype delivered, we will outline the requirements towards realizing the complete vision of our predictive information software product.

(U//FOUO) In this IMC R&D contract we not only leverage the state-of-the art in data modeling, analysis and processing tools, we lead the state-of-the-art with our innovative application of sparse factor-graph models for obtaining reduced complexity representations for large systems, and our proprietary graph learning algorithms for constructing system models even when no prior knowledge of the data probability distribution functions is assumed.

3 (U) Description of Prototype

3.1 (U) Non-technical overview

(U//FOUO) The Product Concept for this IMC BAA is a predictive information software framework capable of automatically learning and drawing inferences upon an arbitrary system of multi-source data. The software will feature our innovative probabilistic models for system representation, learning, and inference processing. We envision our software used as a basic modular component of a broader data discovery, analysis, exploitation and management system, capable of delivering our innovative suite of methods to the challenge problems of the NSA mission. Our complete Prototype will comprise a plurality the multiple disparate data types, *e.g.* structured/unstructured text data and audio/image data–including data translation modules for transforming the data to efficient representations for analysis and processing of a given query. However, in this 3 month IMC program, we restrict the focus to structured text data inputs from disparate category sources in order to make rapid development progress on the underlying graph learning and development framework, which is a key technical contribution of this proposal. The Prototype will feature learned plus modeled statistical descriptors, emphasizing our automatic graph learning methods, and will determine a statistical characterization of dominant system variables relating to a given variable of interest.

(U//FOUO) In this IMC contract, we develop a software Prototype that estimates or predicts the state of hidden variables of a system of data given the state of its observed variables using an automatically learned statistical model for the system. Our proprietary graph learning algorithms use training data, such as historical time-series data pertaining to prediction variables of interest, and graph-based representations of the statistical dependence to other system variables. In addition to innovative methods of empirical model generation, our Prototype system will include limited use of theoretical probability models of data dependence, possibly supplied by the user per application, *e.g.* a Gaussian model for Temperature data with learned mean and variance. To demonstrate the Prototype, we will use historical daily stock price and volume data from NASDAQ and historical daily climate data for major U.S. cities from NOAA. The Prototype software package will learn a non-trivial statistical model for a dependent system of variables based on the historical training data and then it will use the derived model to infer values of hidden variables in test experiments such as prediction of future stock price and city temperature, and reconstruction of missing historical data. In principal, the methods developed in this contract are applicable to arbitrary systems of dependent data with unknown probability distributions, given the availability of adequate test data. The Prototype delivered in this three month IMC contract will be restricted to the following use case: (i) numeric data sets with dependent data (*e.g.* stock and climate data are expected to yield data models with cross-dependencies), (ii) availability of historical training data pertaining to the variables of interest, and (iii) limited inclusion of user supplied probability models for data dependence.

(U//FOUO) Our approach features graph model learning and predictive information processing methods using so-called factor-graphs as a basic modeling framework. Specifically, we use factor-graphs to model system dynamics and their associated message-passing algorithms to estimate unobserved system variables. Our proprietary design approach features scalability to large systems (large number of variables) using sparse graphs to efficiently represent/approximate the dynamics of large systems. We will demonstrate to the Government that the use of sparse representations for modeling probabilistic data systems yields computationally efficient methods for automated system learning and predictive information processing, especially for large systems with many variables.

(U//FOUO) Our algorithms leverage proprietary methods and original research performed by the staff of Aquerre Technologies in addition to state of the art tools from the systems engineering and R&D community. A summary of results of this three month contract will be furnished to the Government along with the software Prototype, including demonstration of predictive performance compared against the Kalman filter as a benchmark technique. In addition, software modules for synthetic test data generation (based on abstract probability models) used in the development effort and simulation of the Prototype will be demonstrated and delivered to the Government.

3.2 (U) Technical details

3.2.1 (U) Formulation of the basic problem and solution approach

(U//FOUO) We first consider the following basic problem: Suppose (i) we have a pair of dependent random variables X and Y, (ii) that X is unobserved and Y is observed, and (iii) they have the joint probability distribution function (p.d.f.) $f(X, Y)$. We wish to predict the value of X given the observation of Y.

(U//FOUO) The joint p.d.f. $f(X, Y)$ contains all of the relevant statistical information pertaining to the system of variables X and Y. It represents the physical law connecting X and Y

and, in general, we must somehow learn or approximate $f(X,Y)$ in order to make predictions on X given Y.

(U//FOUO) We consider applications in which the system p.d.f. is learned based on training data, such as stock market prediction based on historical stock price data. Hence, if (x_1, y_1),. . .,(x_N, y_N), represent a sequence of realizations of the random variables X and Y, we seek to estimate the system p.d.f. from the training data sequence and then use our estimate of the system p.d.f. $f(X,Y)$ to make predictive estimates on X, *e.g.* its conditional expected value given the observation $Y = y_i$:

$$\hat{X} \triangleq E[X|Y = y_i; f = \hat{f}] = \frac{1}{\hat{f}(y_i)} \int x\hat{f}(x, y_i)dx \tag{1}$$

where $\hat{f}(x,y)$ represents the estimate of the system p.d.f. and $\hat{f}(y)$ is the corresponding marginal p.d.f. of Y, given by

$$\hat{f}(y) = \int \hat{f}(x,y)dx. \tag{2}$$

3.2.2 (U) Estimation of the system distribution functions

(U//FOUO) Given the predictor for X defined in Equation (1), it remains only to produce the estimate of the system p.d.f., $\hat{f}$. For this, we use the Fourier series expansion of the system p.d.f. $f(x,y)$ and construct the corresponding Fourier coefficients using empirical estimates of the joint statistical moments of X and Y, as follows. The system p.d.f. may be expressed as:

$$f(x,y) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} c_{kl} \exp\left(j\frac{2\pi}{T}(kx+ly)\right), \; -\frac{T}{2} < x, y < \frac{T}{2} \tag{3}$$

where

$$\begin{aligned} c_{kl} &= \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} f(x,y) \exp\left(-j\frac{2\pi}{T}(kx+ly)\right) dxdy \\ &= E\left[\exp\left(-j\frac{2\pi}{T}(kX+lY)\right)\right], \\ &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \left(-j\frac{2\pi}{T}\right)^{m+n} \frac{k^m}{m!}\frac{l^n}{n!} E[X^m Y^n], \end{aligned} \tag{4}$$

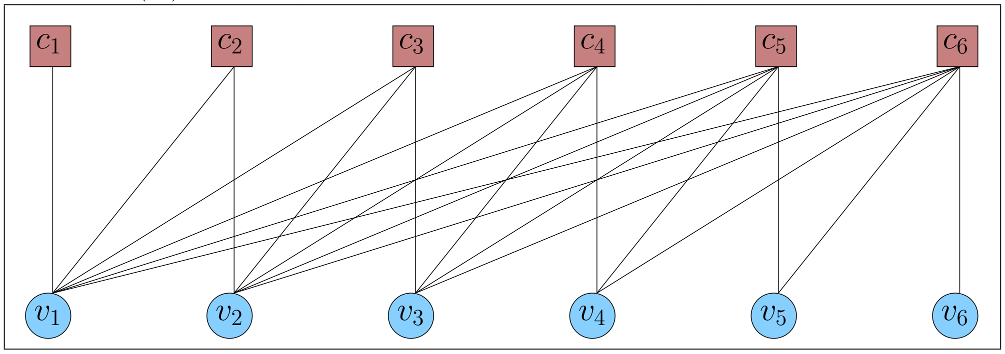
Assuming an stationary system (joint statistics are time-invariant), and independent and identically distributed (i.i.d.) test data (x_i, y_i), $i = 1, \ldots, N$, we can write the following empirical estimates for joint moments of X and Y:

$$E[X^m Y^n] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i^m y_i^n, \; 0 < m, n < \infty \tag{5}$$

Hence, by measuring all the joint moments of X and Y from the test data using (5), we can exactly construct the system p.d.f. using (3) and (4). In practice, reduced complexity estimates can be obtained by measuring only the low order moments (small m and n in Equation (5)).

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Figure 1: (U) Factor graph for six variable system given by the chain rule.



UNCLASSIFIED//FOR OFFICIAL USE ONLY

(U//FOUO) In general, X and Y represent collections of random variables (random vectors) from large data sets. We handle the issue of computational complexity when scaling to large systems using sparse graphs. The graphical modeling framework that we use in this R&D effort is adopted from the coding/information theory community and offers the following reduced complexity features: (i) factorization of the overall system p.d.f. into the product of several less complicated p.d.f.s using the factor graph modeling framework, and (ii) sparse representation methods for modeling sparsely dependent systems and approximating arbitrary physical systems by their dominant statistical relationships.

3.2.3 (U) Factor graphs and the sum-product algorithm

(U//FOUO) The research Prototype delivered to the Government in this IMC contract is based on a factor-graph modeling framework and application of the sum-product algorithm for estimating the state of hidden variables given the state of observed variables (see *e.g.* [7]). Factor-graphs comprise a simple framework for representing an arbitrary hidden/observed variable system, and the widely applicable sum-product algorithm is important due partly to its success in the development of high-performance ECCs.

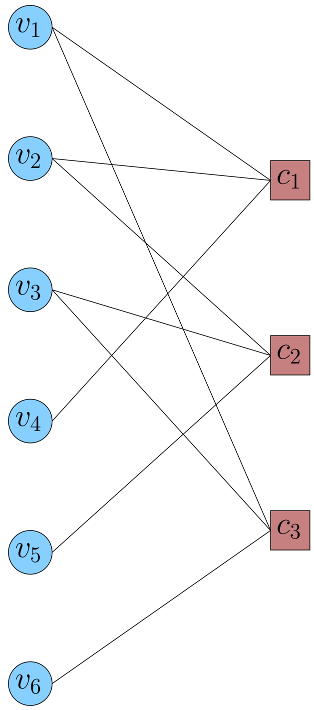
(U//FOUO) A factor-graph is a 'bi-partite' graph, in which there are two-types of nodes and edges connecting only to nodes of different type (sometimes called Tanner graph in coding theory). The 'variable nodes' represent the variables of the system and the 'check nodes' represent the component factors of the overall system p.d.f. (local statistical dependence).

(U//FOUO) An arbitrary system of variables $v_1, \ldots, v_6$, can always be factorized using the chain rule of probability:

$$f(v_1, \ldots, v_6) = f(v_1) f(v_2|v_1) \cdots f(v_6|v_1, \ldots, v_5). \tag{6}$$

Figure 1 depicts the factor-graph representing the chain-rule factorization given in Equation (6). Each system variable is associated with a variable node and each component factor in the factorized system p.d.f. is associated with a check node. Edges of the factor-graph connect check nodes to variable nodes only when the corresponding factor is a function of the corresponding variable. The representation in Figure 1 is generic in the sense that the chain rule factorization applies to any system p.d.f.—we show next how simplified factor-graphs arise from conditional independence of system variables.

Figure 2: (U) Factor graph of the (6,3) Hamming code.



(a) A Tanner graph representing the parity check equations of the (6,3) Hamming code. Edges on the graph are in one-to-one correspondence with 'ones' of the dual matrix, H.

Generator Matrix:

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Input: $\mathbf{u} = [v_1, v_2, v_3]$

Output: $\mathbf{v} = \mathbf{u}G$

Duality Condition: $GH^T = 0$

Dual (Parity-Check) Matrix:

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Valid codeword if and only if:

$\mathbf{v}H^T = \mathbf{u}GH^T = 0.$

(b) Generator and parity check matrices for the (6,3) Hamming code, in systematic form (input-bits are first part of codeword). Any valid codeword must satisfy the parity check equations defined by the dual matrix.

UNCLASSIFIED//FOR OFFICIAL USE ONLY

(U//FOUO) Suppose next that $v_1, \ldots, v_6$ is the codeword of a (6,3) Hamming code in systematic form, where v_1, v_2, v_3 are independently distributed input bits and $v_4 = v_1 + v_2$, $v_5 = v_2 + v_3$, and $v_6 = v_1 + v_3$ are the parity bit equations. Then we can re-write (6) as:

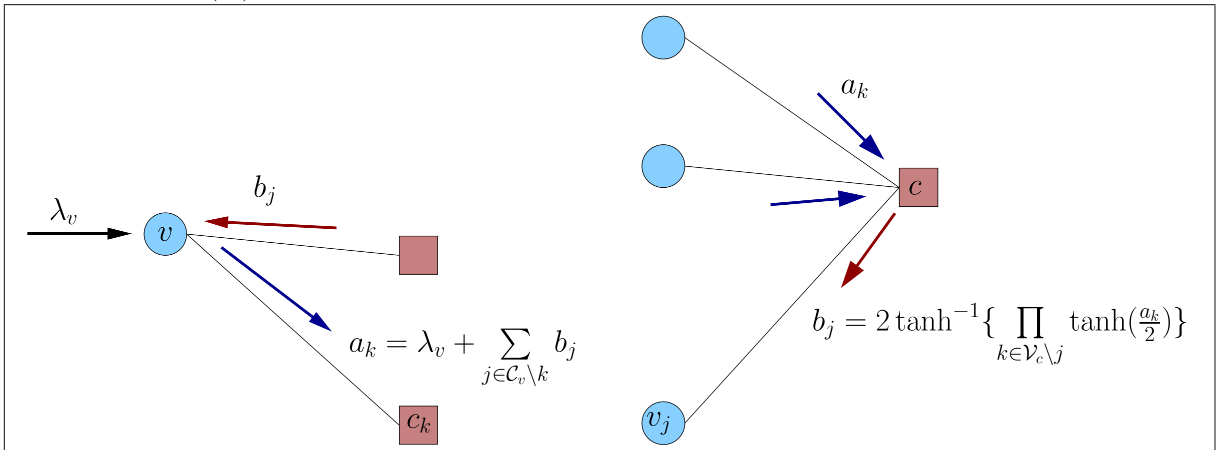
$$f(v_1, \ldots, v_6) = f(v_1)f(v_2)f(v_3)f(v_4|v_1, v_2)f(v_5|v_2, v_3)f(v_6|v_1, v_3), \tag{7}$$

where $f(v_i)$, $i = 1, 2, 3$, represent marginal p.d.f.s (priors) associated with the input bits and the parity bits are distributed according to:

$$f(v_4|v_1, v_2) = \begin{cases} 1, & \text{if } v_4 = v_1 + v_2 \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

etc. Note that $f(v_4|v_1, v_2, v_3) = f(v_4|v_1, v_2)$ is equivalent to the statement that "v_3 and v_4 are conditionally independent given v_1 and v_2". Systems with conditional independence properties among the their variables lead to simplified factor-graph representations and reduced complexity algorithm implementations. Figure 2(a) shows the factor graph representation of the system p.d.f. of the (6,3) Hamming code in Equation (7). The prior distributions on the input bits are assumed to be uniform (therefore contributing a constant factor to the overall p.d.f.) and are omitted from the graph. Figure 2(b) has the mathematical details of the corresponding (6,3) Hamming code, a

Figure 3: (U) Processing steps of the sum-product algorithm with binary data.



(a) Variable node update. $\{\lambda_v\}$ denote prior distributions or observed variable states. $\{a_k\}$ and $\{b_j\}$ are the messages computed by the 'message-passing' algorithm.

(b) Check node update. The component processing steps of the check and variable nodes are performed iteratively until the solution is reached.

UNCLASSIFIED//FOR OFFICIAL USE ONLY

linear block code. The novelty of the algorithms developed in this program stem from exploiting reduced complexity representations resulting from conditional independence of system variables, and from approximating more general and large systems using sparse factor graph models.

(U//FOUO) Given a factor graph model, we can estimate its component p.d.f.s using empirical measurements of relevant system moments derived from test data. In practice, we may only use the low-order moments to generate quick and rough estimates of factors of the system p.d.f. Given the graph model and estimates of its corresponding factors, the sum product algorithm can be directly applied to compute the conditional p.d.f.s of hidden variables of the graph given the status of observed variables. Component processing steps of the sum-product algorithm are illustrated in Figure 3(a) and 3(b), for binary data.

3.2.4 (U) Extension to large systems

(U//FOUO) In principal, the methods outlined in Sections 3.2.1, 3.2.2 can be used to estimate the p.d.f. of an arbitrary system of variables, and Section 3.2.3 outlines graph-based representation methods for algorithm development and implementation, for estimating the state of hidden variables given the state of observed variables. However, the training based approach as described is numerically complex and quickly becomes infeasible as the number of variables gets large. In order to mitigate complexity for large systems (large number of variables), we propose to implement our moment-based estimation technique for determining the local distribution functions of large systems using *sparse* factor-graph models. We define sparse graphs to be factor graphs in which the number of edges per factor remains low even as the number of system variables grows large. In principle, as depicted in Figure 1, a factor of the graph can be connected to every variable of the system, however, in sparse models this level of connectivity is extremely rare and more commonly factors of the model have low degrees, *e.g.* two and three edges (variables) per factor.

An innovation of our proposed approach is restricting the model to sparse representations when scaling to large systems, thereby yielding tractable system representation and statistical learning algorithms. A design challenge of the three month R&D plan will be to develop methods for automatic discovery of good sparse representations based on the available training data for arbitrary numeric systems. By constraining the system model to sparse graphs, we are effectively imposing the constraint that statistical relationships of any system variable are dominated by a few other system variables. This is equivalent to the condition that any given variable is conditionally independent of the remaining system variables when conditioned upon its dominant connected variables.

(U//FOUO) Hence, sparse graphs, such as those used in the design and decoding of Low Density Parity Check (LDPC) codes, maintain a relatively low number of edges even as the number of system variables grows large. Sparse models have factors of small dimension (fewer variables) and require less computational effort to generate moment-based estimates of component factors using training based data. In this IMC contract we extend our proprietary factor graph-based learning approach outlined in Sections 3.2.1, 3.2.2, and 3.2.3 to large systems using sparse graph approximations to system dynamics.

3.3 (U) Prototype data sources

(U//FOUO) We plan to use historical stock price data from Nasdaq.com and historical climate data from NOAA.gov to demonstrate the Prototype developed in this IMC contract. Nasdaq.com has historical daily stock price and volume data for its listed stocks and indexes covering the past ten years. Their web page can be queried using the link in Table 1 and the data is provided as comma separated values. In addition, historical multi-station daily climate data is available for major U.S. cities from NOAA.gov. In order to obtain the data, a web request must be initiated on the NOAA.gov web page and a link to the data download is provided via e-mail. These data sets, selected to demonstrate our Prototype to the Government, represent disparate sources with possible interesting statistical dependence that will be exploited by the predictive information algorithms developed in this IMC contract.

Table 1: (U) Links to proposed data sources.

Data Source	**Description**	**URL**
nasdaq.com	Daily trading data for Nasdaq listed stocks and indexes, for up to 10 years	`http://www.nasdaq.com/quotes/historical-quotes.aspx`
noaa.gov	Multi-station daily historical climate data for U.S. cities	`http://www.ncdc.noaa.gov/cdo-web/search`

UNCLASSIFIED//FOR OFFICIAL USE ONLY

4 (U) Personnel

4.1 (U) Noah B. Jacobsen, P.I.

(U//FOUO) Dr. Jacobsen will be the lead performer on this IMC FY15 contract. He is an expert in field of communications, control and signal processing systems, he has numerous academic publications, industry experience with Alcatel-Lucent Bell Labs, experience as an Adjunct Professor, and is the holder of a U.S. Patent for the invention of a high performance rate-compatible low density parity check code. He is the founder of Aquerre Technologies and he is scheduled for 420 hours over the course of this 3 month contract. His CV is copied below:

(U) Citizenship: (U) U.S. Citizen

(U) Date of Birth: (U) Feb. 15, 1978

(U) Place of Birth: (U) Esopus, New York

(U) Education:

- (U) Ph.D., 2005, Electrical and Computer Engineering, University of California, Santa Barbara.
  - (U) Specialization: Communication, Control and Signal Processing.
- (U) M.S., 2002, Electrical and Computer Engineering, University of California, Santa Barbara.
- (U) B.S., 2000, Electrical Engineering, Cornell University.

(U) Professional Experience:

- (U) Aquerre Technologies LLC: Founder, CEO, Principal Scientist (May 2013 – present).
  - (U) Research and development contracting and consulting services.
- (U) Dex One: Sr. Operations Research Scientist (Sept. 2012 – March 2013).
  - (U) Advertising technology predictive analytics.
- (U) Columbia University: Adjunct Professor (Spring Semester 2012).
  - (U) Linear Systems Theory, Dept. of Electrical Engineering.
- (U) Alcatel-Lucent, Bell Labs: R&D Engineer (July 2006 – Oct. 2011).
  - (U) Error control codes, cooperative relay codes, physical layer communications research, standardization, and development. Includes experience transitioning algorithms to product teams.
- (U) Polytechnic Institute of New York University: Adjunct Professor (Fall Semester 2010).
  - (U) Probability Theory, Dept. of Electrical and Computer Engineering.
- (U) University of California, Santa Barbara: Post-Doctoral Researcher (Sept. 2005 – June 2006).
  - (U) Cognitive radio networks, communication theory.
- (U) Toyon Research Corporation: Consultant (Mar. 2006 – June 2006).

- (U) Cognitive radio networks, communication theory.

(U) Conferences:

- (U) Attendee, 2014 National SBIR/STTR Conference and Short Course "How To Develop An Acceptable Accounting System", Washington, DC, Jun. 16–18, 2014.

(U) Certifications:

- (U) California Basic Educational Skills Test (CBEST) Completed, December 2013.

(U) Patents:

- (U) N. Jacobsen and R. Soni, "Method and system for encoding data using rate-compatible irregular LDPC codes based on edge growth and parity splitting", U.S. Patent No. 7,966,548, June 2007.

(U) Book Chapters:

- (U) G. Barriac, N.B. Jacobsen and U. Madhow. "Chapter 5: The role of feedback, CSI and coherence in MIMO systems," in: Space-time wireless systems: From array processing to MIMO communications, H. Bölcskei, D. Gesbert, C. B. Papadias and A.-J. van der Veen (Ed.), Cambridge University Press, 2006.

(U) Achievements:

- (U) Session Chair, "Wireless Networks and Communications," 43rd Conference on Information Sciences and Systems (CISS), March 18-20, 2009.
- (U) 3GPP2 Ultra Mobile Broadband (UMB) Air Interface Specification: Recognition of Contribution, LDPC Ad Hoc Group, 2007.
- (U) National Science Foundation (NSF) and Japan Society for the Promotion of Science (JSPS) East Asia Summer Institutes Fellowship, Yokohama National University, Japan, 2003.
- (U) Microelectronics Innovation and Computer Research Opportunities Scholarship, University of California, Santa Barbara, 2000–2001.
- (U) Theodore C. Ohart Scholarship in Engineering, Cornell University, 1999–2000.
- (U) Cornell University College of Engineering Cooperative Education Program, with Floyd R. Newman Laboratory of Nuclear Studies, Cornell University, 1998–1999.
- (U) Cornell University Dean of Students Service Award: "Selections Director," Cornell Concert Commission, 1998 and 1999.

(U) Professional Trade Societies:

- (U) Institute of Electronic and Electrical Engineers (IEEE)
- (U) Armed Forces Communications and Electronics Association (AFCEA)

(U) Software and Programming Experience:

- (U) GNU Linux (expert), Debian, C/C++, Bash Shell Scripting, OpenOffice (Writer/ Calc/ Impress), Octave (similar to MatLab), Java, SQL, BASIC, HTML, Emacs, Vi, Gimp, Ardour, Audacity, Mozilla

5 (U) Facilities, Software Frameworks/ Data Sets & GFI

(U//FOUO) All Facilities, Software Frameworks, and Data Sets will be furnished by Aquerre Technologies LLC. The location of performance is 1445 Colby Ave #3, Los Angeles, CA 90025. All software will be developed and provided by Aquerre Technologies, and all 3rd party data will come from open sources. the use of Government Furnished Information (GFI) is not included in this contract.

6 (U) Technical Data or Computer Software

(U//FOUO) The deliverable furnished to the Government in this IMC contract is computer software for graph learning and inference processing with numeric data sets. The Prototype will be demonstrated with data obtained from NASDAQ and NOAA, however the test data used for demonstration purposes is not included as part of the deliverables for this contract. In addition to the test data, the research and development effort will further utilize synthetic data generated from sparse graph system models, and other statistical data models of interest. The software for synthetic data generation will further be included in the contract deliverables.

7 (U) Sub-Contracts or Relevant Collaborations

(U//FOUO) There are no sub-contracts included in this proposal, and Aquerre Technologies LLC (Noah B. Jacobsen) is the sole performer of work on the contract work plan.

8 (U) Other Parties

(U//FOUO) This proposal has not, and will not, be distributed to any other parties, except for the ARC in response to the IMC FY15 BAA.

9 (U) Deliverables and Government Meetings

(U//FOUO) Deliverables shall include monthly project status reports, which shall be submitted via e-mail to the Government, and the computer software product described in Section 9.1. A demonstration of the functional Prototype will be conducted upon completion of the effort at a Government facility within a 25-mile radius of NSA, Fort Meade, Maryland.

9.1 (U) Computer software product

(U//FOUO) *[The following requirements are copied from the BAA-IMC-15 RFI.]* The minimum set of deliverables for the Prototype under the Computer Software Product is:

(U//FOUO) (1) A functional Prototype that, at a minimum, performs the basic functions of the system it is supposed to represent to the degree necessary to clearly and unambiguously

Table 2: (U) Proposed work schedule.

Item	Dates	Description
Part 1	7/1/15-7/31/15	Initial programming and development work, implementing the graph based system model and p.d.f. estimation functions, and development of synthetic data generators. The Part 1 status report (Due: 8/1/15).
Part 2	8/1/15-8/31/15	Intermediate model and methods development. Testing with synthetic data, integration of real data into development Prototype. The Part 2 status report (Due: 9/1/15).
Part 3	9/1/15-9/30/15	Advanced prototype development. Real data testing. Advanced model and methods development. The final report and completed Prototype (Due: 10/1/15).
Demo meeting	TBD: 10/1/15-10/31/15	Schedule prototype demonstration and delivery meeting with Government/Aquerre Technologies during this time window.

UNCLASSIFIED//FOR OFFICIAL USE ONLY

demonstrate the effectiveness of the approach. It is understood that the time and funding constraints of this program may preclude development of a robust and fully operational Prototype but the portions of the system pertaining to the specific technology or approach being proposed as a solution will need to be developed to the extent necessary to clearly show the viability and feasibility of the technology/approach.

(U//FOUO) (2) A description of all simulators, data generators, interface mock-ups, or simplifications that were used in the Prototype because the systems or components they represented were not available for use in the Prototype due to classification issues or other external factors. The intention of this deliverable is to provide comprehensive and full disclosure of which components in the Prototype are fully functional and which components would require additional development prior to a production deployment.

(U//FOUO) (3) List of any required 3rd party support tools or software used in the Prototype or necessary for a potential production deployment of the Prototype.

(U//FOUO) (4) Source code for all non-commercial software used in the Prototype along with the installable executables for the Prototype sufficient to allow the Government to install and demonstrate the Prototype in a Government Research lab.

(U//FOUO) (5) Installation instructions, basic user instructions, and any other materials needed to allow the Government to install and demonstrate the Prototype in a Government Research lab.

(U//FOUO) (6) Software and documentation deliverables shall be provided in soft form on physical media of either data CD or DVD. Flash drives, USB drives, thumb drives, or other types of removable media shall not be used.

9.2 (U) Licensing

(U//FOUO) No third party licenses are required for the products delivered to the Government in this IMC FY15 contract.

References

[1] T. Kailath. *Linear Systems.* Prentice-Hall, 1980.

[2] S. Haykin. *Adaptive Filter Theory.* Prentice Hall, 4th edition, 1982.

[3] A. Papoulis and S.U. Pillai. *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, 4th edition, 2002.

[4] N.B. Jacobsen. *Bayesian approaches to noncoherent communication: From Shannon theory to practical architectures.* PhD thesis, University of California, Santa Barbara, 2005.

[5] J. Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[6] D.J.C. MacKay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, 2003.

[7] F.R. Kschischang, B.J. Frey, and H. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theory*, 47(2):498–519, Feb. 2001.

[8] N.B. Jacobsen and R. Soni. Design of rate-compatible irregular LDPC codes based on edge growth and parity splitting. In *Proc. IEEE Veh. Tech. Conf. (VTC)*, Baltimore, MD, USA, Sept. 2007.

[9] N.B. Jacobsen and R. Soni. Method and system for encoding data using rate-compatible irregular LDPC codes based on edge growth and parity splitting. U.S. patent, No. 7,966,548, June 2011.

UNCLASSIFIED//FOR OFFICIAL USE ONLY