A POLYTHEMATIC REAL-TIME SYNERGISTIC HYBRID DATA TELECOMMUNICATION SYSTEM FOR SCIENTIFIC RESEARCH WITH BIDIRECTIONAL FUZZY FEEDBACK PEER REVIEW BY EXPERT REFEREES

Panagiotis Petratos [1*]

[*1]*Dept of Computing and Information Systems, University of Luton, Park Square, Luton, LU1 3JU*
Email: peter.petratos@luton.ac.uk

ABSTRACT

Heterogeneous research environments, interests and locations do not necessarily coincide, thus hitherto the primary method of communication amongst researchers has been email. In this article a novel unified polythematic, real-time, synergistic, data telecommunication system is proposed with peer-reviewed, bidirectional fuzzy feedback for research scientists, to facilitate scientific information exchange via the extensible markup language (XML) on multiple scientific topics, e.g. in mathematics, physics, biology and chemistry.

Keywords: Data telecommunication system, Scientific research, Bidirectional fuzzy logic

1. Introduction

In the modern, open, heterogeneous research environment, interests and locations do not necessarily form a tautology. Therefore the primary method of communication amongst scholars and researchers at dispersed Universities around the world has hitherto been email. The idea of a disconnected message passing communication can be very effective if scientists are primarily working in isolation and subsequently subject their research work to expert peer review for valuable feedback to improve the quality of their journal publications for instance. However, this process requires research scientists to incubate and refine their ideas primarily in isolation and subsequently articulate them in a formal schema and expose them to expert peer review. If the barrier of initial isolation is surmounted then the process of generating novel, original ideas can be much faster, more effective, fertile and fruitful. This is the aim of the original unified polythematic, real-time, synergistic, data telecommunication system for scientific research, which can be achieved through bidirectional real-time scientific authorship and expert peer-review using a polythematic variety of scientific XML language vocabularies (Murray-Rust & Rzepa, 2002a; Park & Hunting, 2002; Tidwell, 2001; Vlist, 2002), editors and browsers to facilitate the communication of mathematical equations (Sandhu, 2002), chemical molecular compositions (Murray-Rust & Rzepa, 1999; 2002b), protein bond structures (Eisenberg, 2002), and the exchange of scientific data amongst all interested research scholars. Furthermore, the novel data telecommunication system proposed also maintains a history of all the scientific ideas exchanged and is connected to an ever accessible database to which all participating authors may submit research pre-prints. These are subsequently available for bidirectional fuzzy (Petratos & Chen, 2002; Petratos, Chen, Wang & Forsyth, 2002) feedback (including detailed corrections and comments) peer review by expert referees according to their corresponding area of research expertise.

2. System Overview

A representative mathematical model of a digital communication system with a signalling rate 1/T bits/s over a channel of known bandwidth without inter-symbol interference was first introduced by Nyquist as defined by Eq. (1):

$$s(t) = \sum_{n} \alpha_n \cdot g(t - n \cdot T) \qquad (1)$$

where $\alpha_n = \{\pm 1\}$ (Nyquist, 1924). Nyquist determined that for a frequency bandwidth W Hz the maximum bit rate is 2W pulses/s, which permits accurate data recovery and reconstruction at sampling instants k/T, $k = 0, \pm 1, \pm 2, \ldots$ producing the optimum pulse form which is defined by the following equation (2):

$$g(t) = \frac{\sin 2\pi Wt}{2\pi Wt} \tag{2}$$

Hartley, motivated by Nyquist's work, determined that there is a maximum reliable data rate that can be accurately transmitted, with the presence of noise and interference, over a band-limited channel of maximum amplitude $A_{\max}$ and amplitude resolution A_{δ} (Hartley, 1928). Wiener determined the optimum linear filter, which produces the best mean square approximation of the optimum signal waveform s(t) in the presence of additive noise n(t), often encountered in signal demodulation, and results in the received signal (Wiener, 1949), which is defined by Eq. (3):

$$r(t) = s(t) + n(t) \tag{3}$$

Shannon proposed the sampling theorem (Shannon, 1948), which explains that a band-limited W Hz signal can be accurately reconstructed from samples obtained at the Nyquist rate of 2W pulses/s utilizing the interpolation Eq. (4):

$$s(t) = \sum_n s\left(\frac{n}{2W}\right) \cdot \frac{\sin\left[2\pi W\left(t - \frac{n}{2W}\right)\right]}{2\pi W\left(t - \frac{n}{2W}\right)} \tag{4}$$

The difference between analogue and discrete signals is that in the former, x(t) is uniquely defined for all t as a continuous function of time whereas in the latter, $x(k \cdot T)$ exists only at discrete, fixed time intervals T characterized by a sequence of integers k elucidated for each $k \cdot T$ instance. For example, an electrical signal is represented as voltage v(t) or current i(t) with instantaneous power which is defined by Eqs (5) & (6):

$$p(t) = \frac{v^2(t)}{R} \tag{5}$$

$$p(t) = i^2(t) \cdot R \tag{6}$$

across a resistor R which is often normalized to 1 Ω in data telecommunication systems. Therefore the normalized expression of instantaneous power becomes as in Eq. (7):

$$p(t) = x^2(t) \tag{7}$$

where x(t) is either a voltage or a current signal with dissipated energy which is defined by Eq. (8):

$$E_x^T = \int_{-T/2}^{T/2} x^2(t) dt \tag{8}$$

Also, the average dissipated power of x(t) is defined by Eq. (9):

$$P_x^T = \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) dt \tag{9}$$

over the $\left(-\frac{T}{2}, \frac{T}{2}\right)$ interval. Reliable signal transmission and accurate detection and reconstruction depends on the energy level of a signal, which ultimately determines the data telecommunication system's performance. Also, the electromagnetic fields intensities from the waveguides connecting the transmitter to the antenna, as well from the antenna's radiating elements, are analogous to the voltage applied to the transmitter and consequently its power. Hence, for a finite energy signal E_x, where $0 < E_x < \infty$, (Haykin, 1983; Shanmugam, 1979), the following energy Eq. (10) holds true:

$$E_x = \lim_{T \to \infty} \int_{-T/2}^{T/2} x^2(t) dt \tag{10}$$

Also, for a finite power signal P_x, where $0 < P_x < \infty$, (Haykin, 1983; Shanmugam, 1979), the power Eq. (11) holds true:

$$P_x = \ell im_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) dt \tag{11}$$

Shannon continued Hartley's work and adopted a logarithmic measure for the information source's content, demonstrating that the channel capacity C in the presence of additive white Gaussian noise is a function of the bandwidth W, the average transmitter power P and the average noise power N (Shannon, 1948), also known as the Shannon-Hartley capacity theorem which is defined by Eq. (12):

$$C = W \cdot \ell og_2 \left(1 + \frac{P}{N} \right) \tag{12}$$

The average noise power N in (12) is analogous to bandwidth which is defined by the following Eq. (13):

$$N = W \cdot N_0 \tag{13}$$

Hence, if N from (13) is substituted in (12), the latter becomes as in Eq. (14):

$$C = W \cdot \ell og_2 \left(1 + \frac{P}{W \cdot N_0} \right) \tag{14}$$

A digital message is comprised of a stream of bit-collections, each of size k-bits forming a character symbol as defined in (15):

$$m_i = m_1, m_2, \ldots m_M \tag{15}$$

Every character symbol from (15) belongs to a finite alphabet set whose size is defined in Eq. (16):

$$M = 2^k \tag{16}$$

Also, each character symbol from (15) is represented by a corresponding digital waveform as defined by Eq. (17):

$$s_i(t) = s_1(t), s_2(t), \ldots s_M(t) \tag{17}$$

which is transmitted for T seconds, i.e. the symbol duration time, with a data rate which is defined in Eqs (18) & (19):

$$R = \frac{k}{T} \Rightarrow \tag{18}$$

$$R = \left(\frac{1}{T} \right) \cdot \ell og_2 M \tag{19}$$

The adopted code for the proposed hybrid data-telecommunication system encodes a character sequence into binary digits forming a digital bit-stream according to the k = 7, M = 128 encoding schema. The architecture of the proposed hybrid data-telecommunication system builds upon existing standards for internet communication and expands their scope and functionality. In the proposed system architecture the HTTP server is tightly coupled with the telecommunication central switching computing server and together they form a hybrid communication medium, a polythematic hybrid-data telecommunication system. On the internet simply knowing a computer's domain name or internet address is not enough to initiate and complete a transaction since the client computers must also know how to access the specific services required to satisfy the transaction's needs. Hence an automated mechanism is required to distinguish between the influx of received packets and direct the identified information to the appropriate process. This mechanism is implemented through the use of port numbers, which together with the internet address of a computer form a unique computer service address, the contact point for a specific function. Hence ports are the termination points of logical connections that enable long-standing communication sessions for the purpose of providing services to incoming clients. Ports are categorized according to a triad of taxonomies, the well known $2^0 - 1 \cdots 2^{10} - 1$ is used only by processes with root privileges, the registered $2^{10} \cdots 2^{14} + 2^{15} - 1$ is used by ordinary user processes and dynamic or private ports. $2^{14} + 2^{15} \cdots 2^{16} - 1$ is used and assigned by ordinary user processes, unlike the first two categories which are conferred by the assigned internet numbers authority. The following is an architecture diagram of the proposed novel polythematic real-time synergistic hybrid data-telecommunication system.

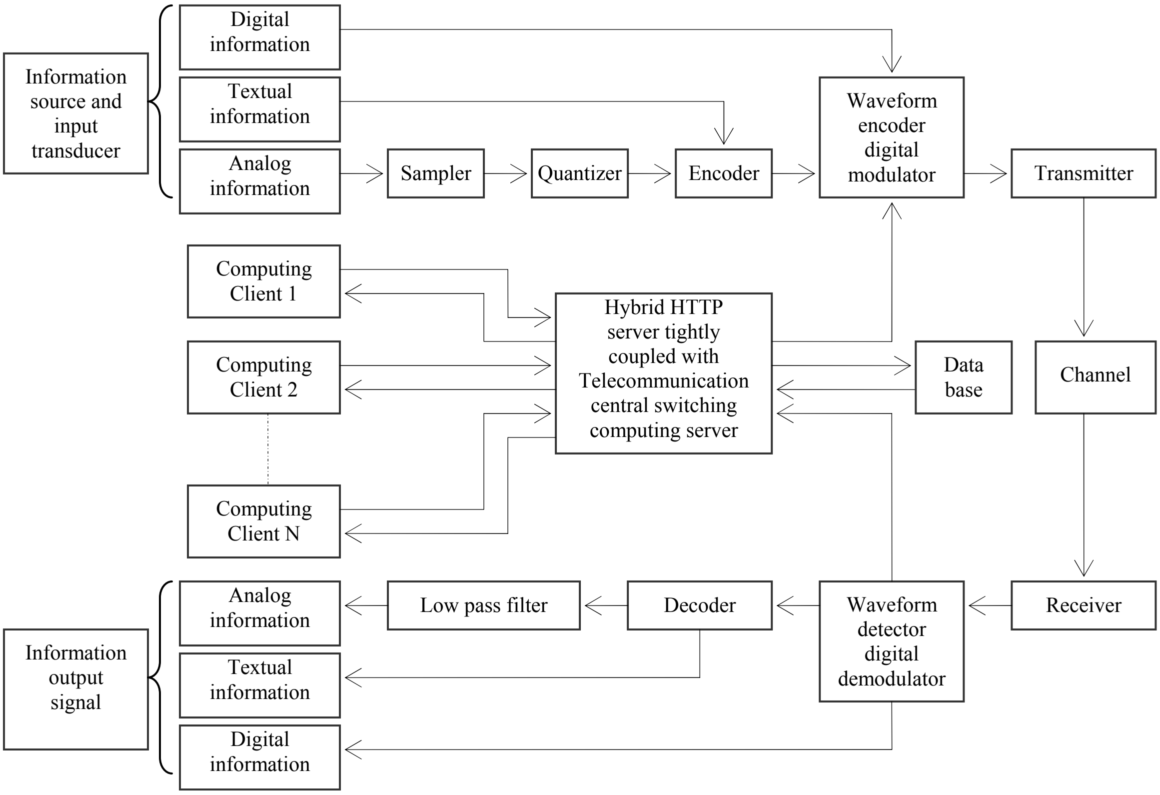

Figure 1. A polythematic real-time hybrid data telecommunication system for scientific research

The proposed novel system supports two modes of communication, a static mode for when clients access information from the database through the well known HTTP port 80 and a dynamic mode. The latter is for when clients access the HTTP server initially, which tightly coupled with the telecommunication server subsequently establishes a dedicated channel for real time communication at port 61969, which is directed and managed only by the telecommunication server. In the architecture diagram the upper logical blocks illustrate the start of the process beginning at the source and moving in the direction of the transmitter signal processing whilst the lower logical blocks illustrate reverse signal processing. The sampler converts the analogue waveform into a discrete-pulse amplitude-modulated waveform. The quantization converts the continuous amplitude waveform samples into a finite set of amplitudes signal. The source encoder converts alphanumeric characters into binary digits and removes redundant information. The modulator converts signals into waveforms compatible with the transmission channel. The transmitter performs a frequency-up conversion and involves a high power amplifier and an antenna while the receiver performs frequency-down conversions and involves an antenna and a low noise amplifier. The demodulator performs signal waveform detection and digital reconstruction. The decoder converts binary digits into alphanumeric characters. Low pass filtering retrieves the original analogue waveform. Data is transmitted by the central-switching computing-server to all other system components as text messages, i.e. character sequences encoded into binary digits forming a digital bit-stream according to the adopted k = 7, M = 128, data-telecommunication system code. For instance, the word *NEW* can be transmitted with the proposed novel system as follows.

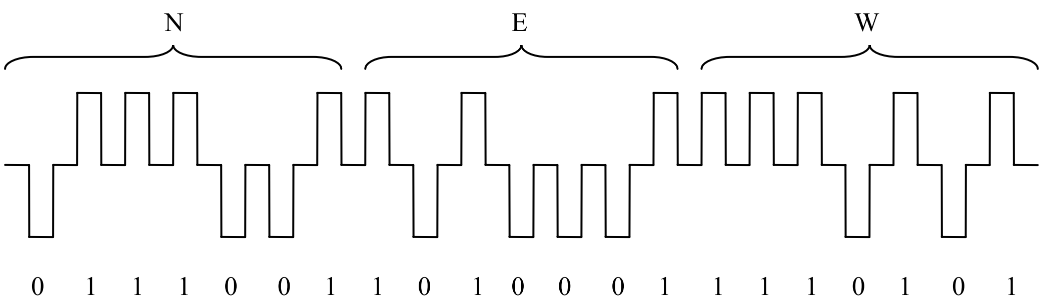

Figure 2. Hybrid data telecommunication system digital bit-stream with the adopted k=7, M=128, code

Each word is broken down into alphanumeric characters which are subsequently encoded into their binary equivalents and transmitted as a digital bit stream, i.e. a digital pulse form consisting of a sequence of ones and zeros. In the above instance the word *NEW* is broken down and each alphanumeric character is encoded into N = 0111001, then E = 1010001, then W = 1110101 according to the adopted k = 7, M = 128, data telecommunication system code. The heart of the hybrid data-telecommunication system is a central-switching computing-server which facilitates real-time communication amongst all system components and maintains a history of the scholarly dialogue in the system's database. Furthermore the central-switching computing-server manages all submitted scientific research pre-prints which are inserted into the system's database in the form of XML documents.

3. A data communication framework based upon polythematic scientific XML languages

A collection of strict syntactic rules for structured data following a hierarchical order, nested inside plain text documents, is the basis of XML (Means & Harold, 2002). Furthermore, XML also forms an extensible framework which allows authors to compose any type of data including their own new specific themed vocabularies that describe for instance, mathematical equations, chemical formulas, and even to graphically represent molecular structural bonds. Thus, XML is a so called meta-language whose principle function is to create other languages, with paired markup tag assertions to define individual character data units called data elements. The robust XML hierarchy defined is clearly evident from the single root element, which contains all other subsequent elements in the document. Hence, the root element is the first element encountered and all other subsequent elements are nested within it, forming a hierarchical order with the root element residing at the base of the tree. This strict hierarchical order allows document authors to create explicit relationships between the data elements of a document. All elements that include other nested elements are called container or parent elements. All nested elements within a parent are called children elements. A parser is special purpose software that processes and verifies the correct syntax of an XML document, making its data elements available to other software applications. In order to prevent naming conflicts between authors who create their own identifiers and between new author-created identifiers and built in class library identifiers, XML provides a facility called namespaces to uniquely identify each element. Namespace prefixes identify the particular namespace to which a unique specific element belongs and differentiates between elements with identical names under different namespaces.

```
<UniversityA:Book
        xmlns:UniversityA = "http://www.UniversityA.edu"
        xmlns:UniversityB = "http://www.UniversityB.edu"
        xmlns:UniversityC = "http://www.UniversityC.edu">
                <UniversityA:Publication> Journal Paper X </UniversityA:Publication>
                <UniversityB:Publication> Journal Paper Y </UniversityB:Publication>
                <UniversityC:Publication> Journal Paper Z </UniversityC:Publication>
</UniversityA:Book>
```

Figure 3. Namespaces assertion example avoiding undesired data element naming collisions

The above example defines the element Publication under the namespaces UniversityA, UniversityB, and UniversityC. In order to ensure the uniqueness of various namespaces, document authors must include universal resource identifiers (URI) for each namespace. Often a URL is assigned as the unique value of a namespace URI because of the unique domain name a URL represents. However, in the previous examples there is no obvious explicitly defined method for displaying or formatting the data elements. An XML vocabulary which addresses this issue is called an extensible style sheet language (XSL) and its subset, which is responsible for creating formatted text documents from XML documents, is called XSL transformations (XSLT). During the transformation process a pair of tree structures are engaged, notably the source tree, which remains unmodified and is the XML document under conversion, and the result tree, which is the computed extensible hypertext markup language (XHTML) document (Tidwell, 2001). The document object model (DOM) also has a hierarchical tree structure that represents each element of the XML document as a node on the tree. Container nodes are called parent nodes while containment nodes are called children nodes, nodes with the same parent are called sibling nodes. An entire node's children and their children and so on are called descendant nodes. Finally the node's parent, and their parents and so on are called ancestor nodes. There is only a single root node called the document root which contains all other descendant nodes. Every node is an object with attributes such as children nodes, tag names, values and methods such as delete, create, append nodes, load XML documents etc. There are two principle methods for parsing XML documents, the event based model which generates events dynamically, invoking listener methods when the parser encounters markup tags as the document is processed, and the DOM which stores the entire document's data elements in a hierarchical tree structure in memory and facilitates fast node access, insertion or deletion. Document type definitions (DTD) and schemas are documents that specify the correct syntax and structure of XML documents by describing which elements are permitted and the attributes each element is allowed to encompass (Vlist, 2002). The parser verifies the XML document structure against the schema and either validates the former as well-formed or issues an error message. However, DTD documents do not adhere to the strict XML hierarchical order; instead they utilize Extended Backus Naur Form (EBNF) grammar to describe data in an XML document. In contrast to DTD's, schemas do not use EBNF grammar; instead they strictly follow the XML hierarchical order, which means that schemas are more extensible and flexible than DTD documents. Therefore an extensible and flexible data-communication framework consists of distributed authored-XML documents, schemas and polythematic scientific XML language vocabularies. The flexibility of XML is evident from the fact that its infrastructure allows authors to create their own novel collection of tags for a thematically unique scientific field of study. These novel collections of tags for describing structured data are called XML language vocabularies. A few examples of the scientific XML language vocabularies include the mathematical markup language (MathML), scalable vector graphics (SVG) and the chemical markup language (CML) (Eisenberg, 2002; Murray-Rust & Rzepa, 2002a, 2002b; Sandhu, 2002). The crucial distinction between the ideas of a scientist and the symbolic mathematical notations that describe them is an important aspect of typography, also known as mathematical typesetting, which has drawn the attention of scientists for a long time (Chaundy, Barrett & Batey, 1954; Higham, 1993; Knuth, 1986; Spivak, 1986; Swanson, 1979; Swanson, 1999). The symbolic notation specifically for mathematics in the XML domain is addressed by MathML. For instance, the following is a MathML matrix example.

```
<mrow><apply><eq/><ci>A</ci>
  <matrix>
    <matrixrow><cn>6</cn><cn>8</cn>
        <cn>4</cn><cn>2</cn></matrixrow>
    <matrixrow><cn><apply><exp/><ci>x</ci></apply>
        </cn><cn><apply><root/><degree><ci>m</ci>
        </degree><apply><plus/><ci>γ</ci><ci>δ</ci>
        </apply></apply></cn><cn>5</cn><cn>9</cn>
    </matrixrow>
    <matrixrow><cn>π</cn><cn><apply><root/><degree>
        <ci>n</ci></degree><ci>ω</ci></apply></cn>
        <cn>λ</cn><cn>μ</cn></matrixrow>
    <matrixrow><cn>φ</cn><cn>ψ</cn><cn>τ</cn><cn>
        <apply><power/><ci>m</ci><cn>n</cn>
        </apply></cn></matrixrow>
  </matrix>
</apply></mrow>
```

$$A = \begin{pmatrix} 6 & 8 & 4 & 2 \\ e^x & \sqrt[m]{\gamma + \delta} & 5 & 9 \\ \pi & \sqrt[n]{\omega} & \lambda & \mu \\ \varphi & \psi & \tau & m^n \end{pmatrix}$$

Figure 4. MathML matrix example, markup is on the left, visualisation is on the right

In the illustrative MathML example above there is a single root data content element, the matrix, which is equal to A and contains four rows of children elements, each, containing another tetrad of elements creating a four by four matrix. The first matrix row child contains four children, a sequence of 4 even numbers, while the second matrix row child contains another four children, exponential to the power of x, the m-root of the sum of variables γ plus δ and a pair of odd numbers. The third and fourth matrix rows each contain another four children. The third matrix row includes 3 variables, π, λ, μ and the n-root of the variable ω. The fourth matrix row includes another 3 variables, φ, ψ, τ and the variable m raised to the power of n, all together to form a square matrix. A triad of categories, notably presentation, content and interface are the taxonomies for all data elements under the MathML language vocabulary. At this point it is worth noting the difference between the content and the presentation data elements. In the example above the matrix data-content element automatically renders opening and closing parentheses and adjusts the matrix size according to the quantity of data encountered in the matrix row data-content elements. In contrast if the presentation data-element *mtable* is used, which may or may not refer to a matrix, the author is responsible for explicitly and manually encoding the opening and closing parentheses according to the quantity of data residing in the table.

CML is another scientific XML language vocabulary which addresses molecular and chemical bond structure representation. For instance, the following is an example of a simple CML molecular bond.

```
<cml:molecule id = “m01” title = “methanol” xmlns:cml = “http://www.xml-cml.org/dtd/cml1_0_1.dtd”>
        <cml:atomArray>
                <cml:atom id = “o1” elementType = “O” hydrogenCount = “1”/>
                <cml:atom id = “c1” elementType = “C” hydrogenCount = “3”/>
        </cml:atomArray>
        <cml:bondArray>
                <cml:bond atomRefs2 = “o1 c1” order = “S”/>
        </cml:bondArray>
</cml:molecule>
```

Figure 5. CML example of a chemical molecular bond structure

In the CML example above there is a single root element, *cml:molecule*, which contains a pair of children elements *cml:atomArray*, *cml:bondArray* and references a unique namespace (Murray-Rust & Rzepa, 1999; 2002a; 2002b). The first child, *cml:atomArray* contains a pair of *cml:atom* children, while *cml:bondArray* contains a single *cml:bond* child. The *cml:molecule* has 3 attributes, id, title and the unique namespace *xmlns:cml*. A priori defined semantics in the namespace attribute ensure that all elements prefixed by *cml:* belong to the unique namespace identified by http://www.xml-cml.org/dtd/cml1_0_1.dtd (Murray-Rust & Rzepa, 1999; 2002a; 2002b) which defines all allowed data element types and their unique meanings and functions, thus avoiding undesirable data-element naming conflicts (Murray-Rust & Rzepa, 1999; 2002a; 2002b). Another scientific XML language vocabulary whose primary function is to describe data-elements representing vector graphics is SVG. The following is an SVG example of a complex molecular structural bond diagram.

```
<svg width=“400” height=“400” xmlns=“http://www.w3.org/2000/svg”>
        <circle cx=“10” cy=“10” r=“20” fill=“red”/>
        <line x1=“10” y1=“30” x2=“50” y2=“100”
        stroke=“black” stroke-width=“1”/>
        <circle cx=“50” cy=“90” r=“20” fill=“pink”/>
        <line x1=“60” y1=“75” x2=“120” y2=“20”
        stroke=“black” stroke-width=“1”/>
        <circle cx=“120” cy=“10” r=“20” fill=“red”/>
        <line x1=“60” y1=“105” x2=“120” y2=“170”
        stroke=“black” stroke-width=“1”/>
        <circle cx=“120” cy=“160” r=“20” fill=“blue”/>
        <line x1=“40” y1=“110” x2=“10” y2=“160”
        stroke=“black” stroke-width=“1”/>
        <circle cx=“10” cy=“160” r=“20” fill=“blue”/>
</svg>
```

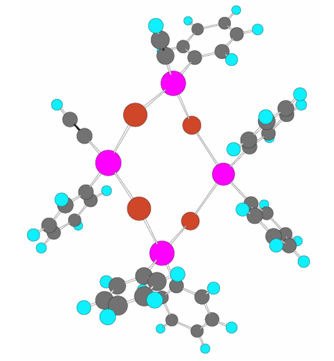

Figure 6. SVG star bond structure, subset markup is on the left, superset complex structural visualisation is on the right.

In the SVG example above the markup on the left describes a sub-molecular star-bond structural diagram that is a small subset of the complex molecular structure diagram displayed on the right (Murray-Rust & Rzepa, 1999; 2002a; 2002b). In the sub-molecular star-bond structural example there is a single root element svg, which contains nine child elements, circles and lines, and references a unique namespace. The pink circle is at the centre of the star-bond structural diagram, the pair of red circles are at the upper left and right extremes and the pair of blue circles are at the lower left and right extremes of the star-bond's structural configuration. The advantages of this approach are quite significant, especially over highly compressed image formats such as jpeg, gif, or png which are currently used for the portrayal and distribution of vector graphics. These highly compressed image formats utilize bitmaps to describe every pixel of an image and utilize various compression algorithms to eradicate pixel data redundancy. This provides space to describe only the more differentiated pixels, which supply the greatest information about how an image is constituted. However the resolution of a bitmap is fixed by its creator and therefore bitmaps cannot scale accordingly when an image is enlarged, which results in poor image quality. In contrast SVG is a text-based format that describes vector graphics in terms of lines, curves and geometric shapes. Therefore, SVG resolution scales accordingly with image size. Furthermore all SVG documents can be searched, dynamically created and automatically computed by machines (Eisenberg, 2002).

4. Strategies for eliciting bidirectional fuzzy refereed expert feedback

In order to ensure high quality journal manuscripts, in the database of the proposed novel data telecommunication system, expert peers acting as referees must review the pre-prints in the repository according to their area of expertise. Therefore their expert feedback must be elicited according to the most desirable strategy. A direct approach in support of scientific collaboration through distributed authorship is proposed by (Gaines, 1993; 1994), whereby group document production for distributed scientific communities across the internet is implemented through the use of digital document archives. Another approach to scientific collaboration involves concept maps that provide a visual representation of knowledge structures and argument forms. Gaines and Shaw (1995a; 1995b) proposed a system supporting collaboration in visual thinking and learning through the distributed authorship over the internet of concept maps representing knowledge. Concept maps support creative processes and are best suited to collaborative learning, particularly in educational groups, by allowing students and teachers to develop concept maps for their domain of interest. By allowing each other to critique, modify or add to the original concept maps they thus gain new knowledge about their subject of interest through the personal views of others (Gaines & Shaw, 1995a, 1995b; Norrie & Gaines, 1995). Eliciting expert feedback about an element's relevance to a collection of elements with a common characteristic, or to an idea or concept, has occupied scholars for a long time. For example, scholars have long been accustomed to constructive criticism by expert peers who act as referees to ensure high quality journal manuscripts of idiosyncratic original research aimed at answering scientifically challenging questions. High quality scientific publications have a common rhetorical modus operandi, notably logos, from the classical triad of rhetorical techniques, the latter pair being pathos and ethos (Parberry, 1994). Not until fairly recently has the topic of refereed expert feedback been revisited in the fuzzy logic field (Klir & Yuan, 1995). Klir & Yuan's proposal for a generalised taxonomy of methods for developing elements' membership functions involving experts' judgments, or feedback, defines two principal categories, direct and indirect methods. In direct methods the questions the experts must answer are rather ambivalent and equivocal; however, they are explicitly connected with the development of the elements' membership function. The direct methods' questions are rather subjective and depend heavily on the experts' idea of how relevant an element is to the concept described by the question. For instance, if an expert is asked to rate the quality of Journal paper X; then the answer would be quite idiosyncratic and not easily quantifiable. Hence, an effective strategy to address such issues is the pair-wise approach categorized under the indirect methods, whereby the expert is not asked directly a question about a single element's degree of membership. Instead a comparison of a pair of elements is requested from the expert, for instance, is Journal paper X of better quality than Journal paper Y (Klir & Yuan, 1995). In indirect methods the questions the experts must answer are less enigmatic and less ambiguous, but are implicitly connected with the development of the elements' membership function. The gnomic questions in direct methods seek expert answers that can attribute a degree of element membership to a linguistic phrase or concept. Furthermore, an alternative strategy to fuzzy logic for acquiring expert knowledge is based on the personal construct theory (Kelly, 1995). An interactive knowledge acquisition system is proposed, for construct elicitation, modelling and comparing personal constructs on the internet, to enable collaborative learners to understand one another's constructs, and gain new knowledge from the diverse richness of various idiographic constructions (Gaines & Shaw, 1993a, 1993b; Shaw & Gaines, 1995). Another popular and very well known method for defining and evaluating the relevance of polar gradient sets of linguistic characterizations to the subject under study is the semantic differential method. Idiosyncratic scientific topics such as appealing design aesthetics, or anti-aesthetics, seismic or volcanic activity, can be characterized by the semantic differential method. Here, each element is represented as a vector with each dimension's value corresponding to a linguistic scale, and the degree of an element's relevance is given by the normalized distance of the vector from the origin of the n-dimensional axis, where n is the number of linguistic characterizations between and including the polar opposites (Osgood, Suci & Tannenbaum, 1957). Furthermore, instead of having just one expert review each paper a more objective approach is to engage several experts' classifications in each paper review. A specific domain of interest is selected along with several experts in the field and the automated system interactively elicits opinions about the papers under study from the experts through natural-language linguistic variables i.e. excellent, very good, good, improvement required, rejected, etc. as well as detailed comments and corrections. These are then expressed in terms of bidirectional fuzzy-element membership functions. In the multiple experts category all the opinions must be computed, hence, if n is the number of experts, x_i is an element of a domain X such as $x_i \in X$, then

$$X = \{x_1, x_2, ..., x_n\} \tag{20}$$

Furthermore, A is a bidirectional fuzzy set representing element membership which is defined in Eq. (21):

$$\mu_A : X \rightarrow [-1.0, \quad 0.0] \cup [0.0, \quad 1.0] \tag{21}$$

If c_i is the competence level of expert i, then the discrete expression of bidirectional fuzzy set A(x) is defined as follows in Eqs (22) & (23):

$$A = c_1 \cdot \mu_A(x_1)/x_1 + c_2 \cdot \mu_A(x_2)/x_2 + \ldots + c_n \cdot \mu_A(x_n)/x_n \tag{22}$$

$$A = \sum_{i=1}^{n} c_i \cdot \mu_A(x_i) \tag{23}$$

If the universe X is an infinite set the continuous expression of bidirectional fuzzy set A on X can be elucidated as follows in Eq. (24):

$$A = \int_X c_i \cdot \mu_A(x_i)/x_i \tag{24}$$

The advantage of the multiple experts approach is that the eventual results are more objective than the single expert approach, especially for ambiguous subjects of study.

5. Experimental data analysis

A series of experiments have been performed on the data-telecommunication system's central-switching computing-server in order to reveal from data analysis the system's behaviour, i.e. response times, request completion rates and the system's capacity, i.e. maximum number of, threads sustained, served and outstanding queued requests, messages sent and received, and maximum percentage of central processing unit utilization. A single thread simulates the actions of a single user recorded earlier and currently executed by the active thread. Therefore, there is an analogy between the number of threads and the number of users. The data illustrated in all graphs are plotted on a logarithmic scale along the y axis so that each data series is clearly distinguishable graphically from the others. The x axis represents data collection points corresponding to 140 second time intervals. Thus, each of the six experiments performed had 121.33 minutes of data collection time. As the experiments progressed the number of threads increased in steady state steps, illustrated in the capacity measurements diagram by the incremental step pulse. The reason for using the steady state intervals was to prevent the system from thrashing by giving it time to adapt and respond reasonably under the increased load.

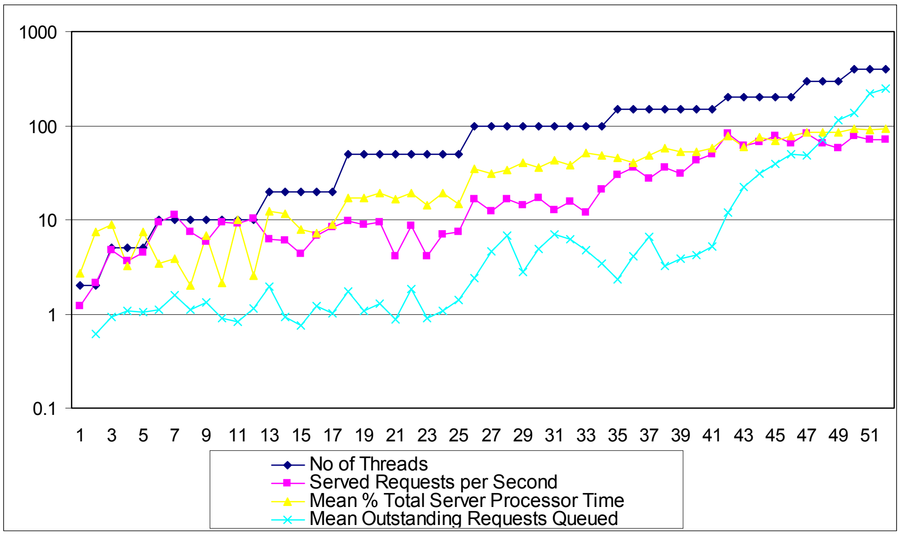

Figure 7. Data telecommunication system capacity measurements

When processes and threads executing on the central-switching computing-server require more pages of memory than there are available in real memory, working memory pages may be swapped out to disk space and then swapped back in when they are needed again. Accessing a page from disk is considerably slower than accessing a page directly from real memory. For this reason, constant page swapping can cause degradation of system performance. Thrashing occurs when the operating system becomes over-committed from continuously swapping pages between the memory and the hard disk so that it does not perform any productive work. Because of competition between the outstanding tasks for real memory the central-switching computing-server can thrash in an effort to complete all the excess current process requests. As the number of simulated users increase, completed requests per second fluctuate with a slight upward inclination following the step-wise thread escalations. At the high end of the served requests per second data series, after the number of threads reached 200, completion rate saturation gradually manifested causing minor vertical fluctuations with an otherwise stable horizontal trend. The mean outstanding requests queued initially are kept to a minimum, which means that the system responds rapidly and completes all the requests in a timely manner. Mean outstanding requests queued fluctuated below ten with a moderately increasing trend until the number of threads reached 200, subsequently causing increased system response delays and a precipitous increase in the mean outstanding requests queued. Mean total server processor time initially fluctuated below ten percent, however as the number of threads increased the central processing unit utilization steadily rose and when the number of threads reached 200, the central processing unit utilization exceeded 80%, gradually entering a steady-system saturation-state. Once the central processing unit's steady saturation-state was attained the whole system became saturated with overcommitted resources and being unable to respond in a timely manner, was overwhelmed by increasing client requests. In this particular case a paired central processing unit configuration would have alleviated the system's saturated state and considerably increased the system's processing threshold.

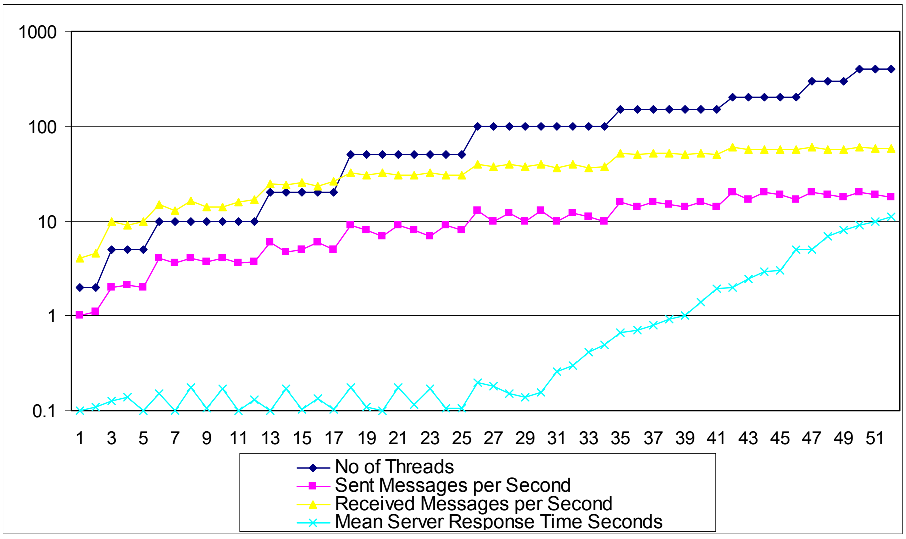

Figure 8. Data telecommunication system behaviour measurements

Initially the mean server response-time was kept to a minimum of less than a quarter of a second which fluctuated with a slight upward inclination. As the number of simulated users increased to 100 the mean system response time steadily rose but still maintained responses below half a second. As the number of threads increased to 150 the mean system response time continued to increase but still maintained a real-time response standard of less than 1.44 seconds. After the number of threads reached 200, system saturation gradually manifested reaching the system's processing threshold at 400 threads and causing a precipitously continuous increase of system response time. When sending messages the transmission control protocol required a greater number of signal processing stages than when receiving messages causing a faster system reception rate and a slower system broadcast rate. Thus, as the number of simulated users increased the received messages per second developed a clear lead over the sent messages per second. As the number of threads increased, sent and received messages per second fluctuated with a slight upward inclination following the step wise thread escalations. At the high end of the sent and received messages per second data series, after the number of threads reached 200, system saturation gradually occurred causing increased system-response time-delays and minor vertical fluctuations with an otherwise stable horizontal trend for the pair of bidirectional transmitted messages per second data series.

6. Conclusion

While email is an adequate method of general communication it is rather disconnected and restrictive, which presents barriers to the initial development of novel research ideas. These barriers can be conquered by a polythematic real-time synergistic research data-telecommunication system that promotes the exchange of ideas with real-time communication while offering a research repository database. This database system provides bidirectional fuzzy-expert refereed feedback whenever the interested scholars are not conversing in real-time. While the system studied reached the processing threshold gradually after 200 threads, the overall response times and system processing throughput was quite satisfactory in the earlier stages. One solution for increasing the system processing threshold is a paired central processing unit configuration, which typically increases the overall data telecommunication system processing threshold by about one and a half times.

7. References

Chaundy, T. W., Barrett, P. R. & Batey, C. (1954) *The Printing of Mathematics, Aids for authors and editors and rules for compositors and readers at the University Press, Oxford*, London: Oxford University Press.

Eisenberg, D. (2002) *SVG Essentials*, Cambridge: O’Reilly and Associates.

Gaines, B. R. (1993) An Agenda for Digital Journals: The Socio-Technical Infrastructure of Knowledge Dissemination. *Journal of Organizational Computing 3*(2), 135-193.

Gaines, B. R. (1994) Supporting Collaboration through Multimedia Digital Archives. *Canadian Multimedia Conference Proceedings*, Alberta, Canada.

Gaines, B. R. & Shaw, M. L. G. (1993a) Eliciting Knowledge and Transferring it Effectively to a Knowledge-Based System. *IEEE Transactions on Knowledge and Data Engineering 5*(1), 4-14.

Gaines, B. R. & Shaw, M. L. G. (1993b) Knowledge Acquisition Tools based on Personal Construct Psychology. *Knowledge Engineering Review 8*(1), 49-85.

Gaines, B. R. & Shaw, M. L. G. (1995a) Collaboration through concept maps. *Proceedings of Computer Supported Cooperative Learning*, Bloomington, USA.

Gaines, B. R. & Shaw, M. L. G. (1995b) WebMap: Concept Mapping on the Web. *Proceedings of WWW4: Fourth International World Wide Web Conference,* Boston, USA.

Hartley, R. (1928) Transmission of Information. *Bell Systems Tech. Journal 7*, 535.

Haykin, S. (1983) *Communication Systems*, New York: Wiley and Sons.

Higham, N. J. (1993) *Handbook of writing for the mathematical sciences*, Philadelphia: Society for Industrial and Applied Mathematics.

Kelly, G. A (1995) *The Psychology of Personal Constructs*, New York: Norton.

Klir, G. J. & Yuan, B. (1995) *Fuzzy Sets and Fuzzy Logic*, New Jersey: Prentice Hall.

Knuth, D. E. (1986) *The Texbook*, Reading: Addison Wesley.

Means, S. & Harold, E. R. (2002) *XML in a nutshell*, Cambridge: O’Reilly and Associates.

Murray-Rust, P. & Rzepa, H. S. (1999) Chemical markup Language and XML Part I. Basic principles. *Journal Chem. Inf. Comp. Sci. 39*, 928.

Murray-Rust, P. & Rzepa, H. S. (2002a) Scientific publications in XML - towards a global knowledge base. *Data Science Journal 1*, 84-98.

Murray-Rust, P. & Rzepa, H. S. (2002b) Markup Languages- How to structure chemistry related documents. *Chemistry Intl. 24*(4), 9-13.

Norrie, D. H. & Gaines, B. R. (1995) The Learning Web: A System View and an Agent-Oriented Model. *International Journal of Educational Telecommunications 1*(1), 23-41.

Nyquist, H. (1924) Certain Factors Affecting Telegraph Speed. *Bell Systems Tech. Journal 3*, 324.

Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. (1957) *The Measurement of Meaning*, Urbana: University of Illinois Press.

Parberry, I. (1994) A Guide for New Referees in Theoretical Computer Science. *Information and Computation 112*(1), 96-116.

Park, J. & Hunting, S. (2002) *XML Topic Maps*, Reading: Addison Wesley.

Petratos, P., Chen, L., Wang, P. & Forsyth, R. (2002) A Bi-directional Fuzzy Logic Theory: The Generalized Knuth's Triadic Logic for Information Retrieval. *Proceedings of the IEEE Systems Man and Cybernetics Conference*, Hammamet, Tunisia.

Petratos, P. & Chen, L. (2002) A note on bidirectional fuzzy logic. *Proceedings of the North American Fuzzy Information Processing Society Conference*, New Orleans, USA.

Sandhu, P. (2002) *The Mathml Handbook*, Hingham: Charles River Media.

Shanmugam, K. (1979) *Digital and Analog Communication Systems*, New York: Wiley and Sons.

Shannon, C. E. (1948) A Mathematical Theory of Communication. *Bell Systems Tech. Journal 27*, 379-423.

Shaw, M. L. G. & Gaines, B. R. (1995) Comparing Constructions through the Web. *Proceedings of Computer Supported Cooperative Learning*, Bloomington, USA.

Spivak, M. D. (1986) *The Joy of TEX A gourmet guide to typesetting with the AMS-TEX macro package*, Providence: American Mathematical Society.

Swanson, E. (1999) *Mathematics into type: Updated Edition*, Providence: American Mathematical Society.

Swanson, E. (1979) *Mathematics into type: Copy editing and proofreading of mathematics for editorial assistants and authors*, Providence: American Mathematical Society.

Tidwell, D. (2001) *XSLT*, Cambridge: O'Reilly and Associates.

Vlist, E. (2002) *XML Schema*, Cambridge: O'Reilly and Associates.

Wiener, N. (1949) *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, New York: Wiley and Sons.