COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE/If not in response to a program announcement/solicitation enter NSF 04-2

NSF 04-561 06/14/04

FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.)

DMS - APPROACHES TO COMBAT TERRORISM

FOR NSF USE ONLY

NSF PROPOSAL NUMBER

0442154

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)

SHOW PREVIOUS AWARD NO. IF THIS IS
☐ A RENEWAL
☐ AN ACCOMPLISHMENT-BASED RENEWAL

IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES ☐ NO ☒ IF YES, LIST ACRONYM(S)

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE

Rensselaer Polytechnic Institute

AWARDEE ORGANIZATION CODE (IF KNOWN)

0028035000

ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE

Rensselaer Polytechnic Institute
110 8th Street
Troy, NY. 121803522

NAME OF PERFORMING ORGANIZATION, IF DIFFERENT FROM ABOVE

PERFORMING ORGANIZATION CODE (IF KNOWN)

ADDRESS OF PERFORMING ORGANIZATION, IF DIFFERENT, INCLUDING 9 DIGIT ZIP CODE

IS AWARDEE ORGANIZATION (Check All That Apply) (See GPG II.C For Definitions)
☐ SMALL BUSINESS ☐ MINORITY BUSINESS
☐ FOR-PROFIT ORGANIZATION ☐ WOMAN-OWNED BUSINESS

☐ IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE

TITLE OF PROPOSED PROJECT **Surveillance, Analysis and Modeling of Chatroom Communities**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ 157,673	12 months	01/01/05	

CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW

☐ BEGINNING INVESTIGATOR (GPG I.A)
☐ DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C)
☐ PROPRIETARY & PRIVILEGED INFORMATION (GPG I.B, II.C.1.d)
☐ HISTORIC PLACES (GPG II.C.2.j)
☐ SMALL GRANT FOR EXPLOR. RESEARCH (SGER) (GPG II.D.1)
☐ VERTEBRATE ANIMALS (GPG II.D.5) IACUC App. Date _____
☐ HUMAN SUBJECTS (GPG II.D.6)
Exemption Subsection _____ or IRB App. Date _____
☐ INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED (GPG II.C.2.g.(iv).(c))
_____
☐ HIGH RESOLUTION GRAPHICS/OTHER GRAPHICS WHERE EXACT COLOR REPRESENTATION IS REQUIRED FOR PROPER INTERPRETATION (GPG I.E.1)

PI/PD DEPARTMENT
Computer Science

PI/PD FAX NUMBER
518-276-4033

PI/PD POSTAL ADDRESS
110 8th Street
Troy, NY 12180
United States

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Electronic Mail Address
PI/PD NAME Bulent Yener	PhD	1994	518-276-6907	yener@cs.rpi.edu
CO-PI/PD Mukkai S Krishnamoorthy	PhD	1976	518-276-6911	moorthy@cs.rpi.edu
CO-PI/PD				
CO-PI/PD				
CO-PI/PD				

Electronic Signature

CERTIFICATION PAGE

Certification for Authorized Organizational Representative or Individual Applicant:

By signing and submitting this proposal, the individual applicant or the authorized official of the applicant institution is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding debarment and suspension, drug-free workplace, and lobbying activities (see below), as set forth in Grant Proposal Guide (GPG), NSF 04-2. Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, Section 1001).

In addition, if the applicant institution employs more than fifty persons, the authorized official of the applicant institution is certifying that the institution has implemented a written and enforced conflict of interest policy that is consistent with the provisions of Grant Policy Manual Section 510; that to the best of his/her knowledge, all financial disclosures required by that conflict of interest policy have been made; and that all identified conflicts of interest will have been satisfactorily managed, reduced or eliminated prior to the institution's expenditure of any funds under the award, in accordance with the institution's conflict of interest policy. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Appendix C of the Grant Proposal Guide.

Debarment and Suspension Certification

(If answer "yes", please provide explanation.)

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency? Yes ☐ No ☒

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Appendix D of the Grant Proposal Guide.

Certification Regarding Lobbying

This certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding $100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding $150,000.

Certification for Contracts, Grants, Loans and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

(1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, Title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than $10,000 and not more than $100,000 for each such failure.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME Richard E Scammell	Electronic Signature	Jun 14 2004 11:22AM

TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS	FAX NUMBER
518-276-6283	scammr@rpi.edu	518-276-4820

*SUBMISSION OF SOCIAL SECURITY NUMBERS IS VOLUNTARY AND WILL NOT AFFECT THE ORGANIZATION'S ELIGIBILITY FOR AN AWARD. HOWEVER, THEY ARE AN INTEGRAL PART OF THE INFORMATION SYSTEM AND ASSIST IN PROCESSING THE PROPOSAL. SSN SOLICITED UNDER NSF ACT OF 1950, AS AMENDED.

Surveillance, Analysis and Modeling of Chatroom Communities

Project Summary

The aim of this proposal is to develop new techniques for information gathering, analysis and modeling of chatroom communications. Internet chatrooms have become common means of interaction and communications, and they carry valuable information about formal or ad-hoc formation of cyber groups with diverse objectives. While the structure of other forms of cyber communications such as email interactions, weblogs, news group posting can be represented by graph-based approaches, elusiveness of chatroom interactions requires new techniques.

Graph-based models fail when there is a lack of a precise "edge" definition induced by a binary relationship between a pair of participants (vertices). Hypergraph-based approaches also fall short in which hyperedges are defined over a subset of vertices. The weight of an hyperedge is the weight of the associated vertices. If these vertices do not have identical weights then the hyperedge weights are not well defined. Since in chatroom communications we expect the hyperedges to have different weights, we cannot model them by hypergraphs.

Intellectual merits of this project are twofold. First, we investigate graph-less models to capture the structure of chatroom communications. In particular we investigate how to develop a multidimensional singular value decomposition (SVD) approach for component analysis of chatroom communication data. Second, we develop new visualization techniques that do not rely on a graph based representation of the data to display the structural information found in the first step. In particular we investigate novel visualization algorithms based on Venn Diagrams to display the structural properties of chatroom communications.

Broad impacts of the project is to provide Intelligence Community (IC) with new tools and techniques for automatic information gathering, analysis and prediction to answer several crucial questions including (i) *in which chatrooms topic A is discussed*, (ii) *who is chatting about topic A in chatroom X*, (iii) is topic A is a hot one in chatroom X etc.

Education: The work will provide research opportunities for graduate students from diverse sub-fields to work together to address the multidimensional structure analysis and its visualization.

Diversity: The potential for students from diverse backgrounds to participate in the research will increase our ability to attract and retain students from under-represented groups. For instance, Dr. Yener is currently supervising two female PhD students one of whom will be focusing her PhD thesis work on this project.

TABLE OF CONTENTS

For font size and page formatting specifications, see GPG section II.C.

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

1 Introduction

Internet chatrooms provide for an interactive and public forum of communication for participants with diverse objectives. Two properties of chatrooms make them particularly vulnerable for exploitation by malicious parties. First, the real identity of participants are decoupled from their chatroom nicknames. Second, multiple threads of communication can co-exist. Although human-monitoring of each chatroom to determine *who-is-chatting-with-whom* is possible, it is very time consuming hence not scalable. Thus, it is very easy to conceal malicious behavior in Internet chatrooms and use them for covert communications (e.g., adversary using a teenager chatroom to plan a terrorist act).

In this work, we consider a fully automated surveillance system for data collection and analysis in Internet chatrooms to discover hidden groups. We propose a system to be deployed in the background of any chatroom as a *silent listener* for eavesdropping. The surveillance is done in the form of statistical profiling for a particular chatter, a group of chatters or for the entire chatroom. Algorithms are designed to use the statistical profiles to determine chatters and their partners and answer to queries including (i) *in which chatrooms topic A is discussed*, (ii) *who is chatting about topic A in chatroom X*, (iii) is topic A is a hot one in chatroom X etc. Thus, the proposed system could aid the intelligence community to discover hidden communities and communication patterns in chatrooms without human intervention. Our previous work [3] indicates that the chatroom communications are quite elusive which presents difficulties for graph-based [5, 2] and SVD based approaches [3]. Thus, this work considers generalization of SVD to higher order tensors to cope with the multiple dimensions of chatroom communications.

Problems with Graph-based Methods: Edges and Hyperedges

In graph-based approaches an edge between i and j represents a binary relationship. However, in chatroom communications establishing such binary relationships may not be possible due to several reasons. First, unless a semantic analysis is performed, it is not obvious if such binary relationship exists between i and j. Second, if a subset of nodes are communicating on the same topic, then a simple graph-based approach would result in cliques which may not be accurate. Similarly, hypergraph-based approaches also fall short in which hyperedges are defined over subset of vertices. The weight of an hyperedge is the weight of the associated vertices. If these vertices do not have identical weights then the hyperedge weights are not well defined. Since in chatroom communications we expect the hyperedges to have different weights, we cannot model them by hypergraphs.

Background and Related Work

IRC (Internet Relay Chat) [8, 9, 10, 11, 7] is the original and the most widely used Internet chat medium. Currently there are *675* IRC networks distributed all around the world. There are total of *5290* servers within these networks having *1219906* users on *591257* channels [6]. IRC is a multi-user, multi-channel and multi-server chat system which runs on a Network. It is a protocol for text based conferencing and allows people all over the world to talk to one another in real time. Conversation or chat takes place either in private or on a public channel called as chat room. There are several spy tools to listen to IRC channels. PieSpy [19, 20] is an IRC program (bot) that logins to IRC servers, joins and listens to IRC Channels. PieSpy is used to collect messages and extract information to visualize social networks in a chatroom. It has simple set of heuristic rules to decide who is talking to who. *Chat Circle* [25] is an abstract graphical interface for synchronous conversation which aims to create a richer environment for online discussions. This approach is based on graphical visualization and does not provide eavesdropping capability. These approaches lack

a solid theoretical foundation and operate with heuristic rules. There are several proposals for discovering hidden groups and communication patterns in social networks (see [15, 5, 2] and references therein). For example Social Network Analysis (SNA) [15] is a software tool which considers relationships between people, groups, organizations, computers or other information/knowledge processing entities. The nodes in the network are the people and groups while the links are the relationships between the nodes. Since building a precise graph-based modeling of chatrooms is not feasible, these approaches are not applicable.

Our first contribution is based on using the multidimensional SVD (i.e., tensors with order higher than two) to capture the multiple facets of the chatroom communications. Multidimensional SVD has been a focus of intensive research [16, 17, 18, 27, 13, 12, 14]. It is well understood that there is no "best" way to generalize SVD to higher dimensions. The computational methods favor greedy algorithms based on iterative computations such as *alternating least squares* (ALS) [27, 18, 16]. While the special cases (e.g., for tensors with orthogonal decompositions [27]) are possible, in general enforcement of constraints in ALS remains a challenge.

Our second contribution is to develop new visualization techniques for multi-dimensional SVD based on Venn Diagrams. Visualization has been used extensively to understand the behavior of underlying systems. Venn diagrams are a type of visualization tool used to depict sets and membership in sets. We will use Venn diagrams to visualize each of the four dimensions. We plan to visualize the evolution of chat room communities using the online visualization of evolutionary Venn diagrams. Visualization of Venn diagrams has been studied in the past by Ruskey [23]. Online visualization of graphs has been studied by Morrison et al [26].

Intellectual merit: First, we investigate graph-less models to capture the structure of chatroom communications. In particular we investigate how to develop a multidimensional singular value decomposition (SVD) approach for component analysis of chatroom communication data. Second, we develop new visualization techniques to display the structural information found in the first step.

Broad impacts: This project aims to provide Intelligence Community (IC) with new tools and techniques for automatic information gathering, analysis and prediction to answer several crucial questions including (i) *in which chatrooms topic A is discussed*, (ii) *who is chatting about topic A in chatroom X*, (iii) is topic A is a hot one in chatroom X etc. This information can be quite valuable to IC since the Internet public forums has become a preferred medium for illegal and adversarial communications replacing traditional mediums such as prepaid cell phones [21].

Education: The research will provide for research opportunities for graduate students from diverse subfields to work together to address the multidimensional structure analysis and its visualization.

Diversity: The potential for students from diverse backgrounds to participate in the research will increase our ability to attract and retain students from under-represented groups. For instance, Dr. Yener is currently supervising two female PhD students one of whom will be focusing her PhD thesis work on this project. We are confident that with strong institutional emphasis by RPI on encouraging students from diverse backgrounds to enter research, together with the attractive projects that the proposed infrastructure will support, we will be able to further attract and nurture many more students. We will also do targeted advertising to minority professional societies such as the Association for Women in Science (www.awis.org), the Minority Science Network (www.nextwave.sciencemag.miscinet), the American Indian Science and Engineering Society (www.aieses.org), and the National Society for Black Engineers (www.nsbe.org).

2 Overview of Proposed Work

2.1 Problem Statement and Contributions

In this project we consider multidimensional SVD generalization with four dimensions: (1) users (who), (2) time (when), (3) topic (what), (4) location (where). Our approach is two pronged.

We will examine how to construct the chatroom tensor so that it has complete orthogonal decomposition. This approach will require careful data analysis to construct the data for each dimension. If a completely orthogonal decomposition of chatroom tensor is possible then we can simply apply the algorithm in [27] with linear time convergence. However, if no such construction exists then we will investigate how to adapt the more general ALS methods [16] to chatroom tensors. In particular a convergence study of ALS methods in the context of chatroom tensors will be conducted. Furthermore, since orthogonal rank decompositions (ORD) are not unique [13], we will examine the impacts of different ORD on convergence of ALS.

2.2 Dimensions of the Chatroom Data Tensor

In this section we will introduce different dimensions of the chatroom communications data.

2.2.1 Chatter IDs

IRC follows a client/server model. Server can be running on many machines, in a distributed fashion on a network, where each server has a copy of the global state information of the entire network. IRC client communicates with an IRC server through Internet. Client can choose the server to log on to it, picks a unique nickname and selects one or more channels to join. A channel is a group of one or more users who receive all messages addressed to that channel. A channel is characterized by its name, topic, set of rules and current members. As a result a user can be active in multiple channels on the same server, in addition to being active on multiple servers.

Multiple nick names further complicate the identification of users while analyzing chat room data. During an IRC session a client can change its nickname on the flight. Although such changes are visible from *change nickname* messages, once clients are disconnected, they can use any nicknames they like when they login to server again. Furthermore, a client can access IRC from different hosts under different nicknames. In such situations, it is impossible to decide whether given two nicknames belong to same or different person. The multiple nicknames problem is out of the scope of this proposal but it certainly introduces inaccuracy to graph based modeling. Thus construction of the first dimension chatroom tensor is a very challenging task with many degrees of freedom.

2.2.2 Chat Times

Our approach is based on silently listening to the channel and logging all the messages. There are several degrees of freedom to construct the time dimension of the chatroom tensor. For example in [3] we used ten days of channel logs for the statistical analysis. For our algorithm, we used two busiest four-hour periods in a day (i.e., from 08:00 to 12:00, and from 20:00 to 24:00). Each four hour period is divided into eight 30-minute intervals called *sampling windows* (e.g., 08:00 to 08:30, 08:30 to 09:00, etc).We also used other sampling window sizes, i.e. $15, 30, 45, 60, 75, 90$-min. We further divide the sampling windows into *time slots* of sizes 15, 30, 45, 60, 75 or 90 seconds.

We use two methods to generate time dimension of the tensor. In *fixed window* in which the sampling windows do not overlap (e.g., 08:00 to 08:30, 08:30 to 09:00, etc). However, there is a chopping of the

conversation at the allotted time interval with the fixed window scheme. Thus, we consider *sliding window* method in which the windows overlap. For example, in sliding window method for window overlap of 15 minutes, we generate matrices for time intervals of 08:00-08:30, 08:15-08:45, 08:30-09:00, etc. This method generates sixteen matrices for 4 hour period for window size of 30 minutes.

2.2.3 Chat Topics- Key Words

Another dimension of the problem is the subject matter being discussed in chatrooms. For example consider a chatroom that focuses on *Philosophy*. There are on-line resources (http://www.glossarist.com/glossaries/humanities-social-sciences/philosophy.asp) providing for a philosophy dictionary, glossary and terms directory. The keywords used by chatters is matched to such the on-line directory to conjecture what topic is being discussed. A directory of keywords and terms can be constructed for any particular topic and chatrooms can be searched to cluster the chatters around topics.

2.2.4 Chatter Location

IRC Servers keep track of the IP or the hostname of the machine from which a nickname (chatter) is connected. When a client connects to an IRC Server, server sends series of DNS and Ident queries [11] to figure out hostname of the host and the account name in the host from which the client initiates his/her connection . If the queries are not successful, IRC Server simply stores source IP address of the TCP connection through which client is connected. IRC Client Protocol [10] provides several commands to learn source IP or hostname of a given nickname: (i) WHO, (ii) WHOIS, (iii) WHOWAS, (IV) USERHOST.

Once the IP address of the client retrieved from an IRC server, it is easy to locate the BGP Autonomous System to which this IP belongs to. We can find the exact geographic location of each Autonomous Systems by using the Latitude and Longitude value obtained from [1] Latitude and Longitude values can be converted to rectangular coordinates and therefore can be located on a rectangular map. Furthermore, [1] also provides tools to plot geographical maps of the Internet. Plot-latlong is a small command line tool written in perl [22] for plotting points on geographic maps given a list of latitude/longitude pairs [1].

It is clear from the above discussion that construction of a chatroom tensor has many degrees of freedom. The immediate goal is to check if the tensor can be constructed in such a way that it has a complete orthogonal decomposition. Different constructions will have significant impact on the computational algorithms and will be studied in the proposed research.

2.3 Visualization Research

As the notion of an "edge" or "hyperedge" is hard to define here, visualization techniques based on graphs as well as hypergraphs will not be applicable here. However, Venn diagrams have been used effectively to represent sets and how different sets have common memberships.

Drawing Venn diagrams become complicated when there are more than six sets. Since each set of Venn diagrams contains information about a dimension, it is likely that we will have more than six sets. We will develop a method to visualize the different dimensions using Venn diagrams.

Chat room communications are continuous, and the communities will be evolving in time. We will try to visualize this evolution through Venn diagrams. The evolution will be animated, based on the partial computational results at different times. Because the SVD computation is expensive, the task will be how to judiciously interpolate the computations.

[1] However there are some challenges: 1-) Client may not be using direct connection, instead: 1.1-) coming through a proxy server 1.2-) coming through a compromised or hacked server 2-) Client might be using an hijacked TCP session (Man-in-the-middle attack result) 3-) Client might be using a public kiosks.

Another important research task is to input semantics to the Venn diagram by associating information on various dimensions. We will add a query interface to the Venn diagram which will not only answer simple questions but also answer queries which can be reasoned using a small number of inferences.

2.4 Results from Prior NSF Support

Dr. Yener and Dr. Krishnamoorthy are Co-PIs on NSF ITR Award #0324947. This research focus on identifying hidden groups in social networks. The work examined the structure of Internet chatrooms by applying SVD on three different chatrooms. The matrices for the SVD algorithm are constructed by using statistical properties of the chatroom communications [3]. This study led to the question of whether using tensors of order higher than two may provide for better solutions without using graph-based approach. Under the same grant we also investigated the structure of email communication graphs again using SVD [4]. We constructed graphs and studied its spectral analysis as well as several topological properties to show that principal component analysis (PCA) can be deployed for data mining in email communications.

References

[1] www.caida.org : http://www.caida.org. (accessed 6 June 2004).

[2] Goldberg, M., Horn, P., Magdon-Ismail, M., Riposo, J., Siebecker, D., Wallace, W., Yener, B.: Statistical modeling of social groups on communication networks. In: First conference of the North American Association for Computational Social and Organizational Science (CASOS 03), Pittsburgh PA, CASOS (2003).

[3] Camtepe, A., Krishnamoorthy, M., Yener B.: A Tool for Internet Chatroom Surveillance. In: Symposium on Intelligence and Security Informatics (ISI'04) June (2004).

[4] Drineas, P., Krishnamoorthy, M., Sofka, D. M., Yener, B.: Studying E-mail Graphs for Intelligence Monitoring and Analysis in the Absence of Semantic Information. In: Symposium on Intelligence and Security Informatics (ISI'04) June (2004).

[5] Magdon-Ismail, M., Goldberg, M., Siebecker, D., Wallace, W.: Locating hidden groups in communication networks using hidden markov models. In: NSF/NIJ Symposium on Intelligence and Security Informatics (ISI 03), Tucson, PA, ISI (2003).

[6] Gelhausen, A.: IRC statistics. http://irc.netsplit.de (1998) (accessed 10 February 2004).

[7] Johns, M.S.: RFC 1413 Identification Protocol (1993).

[8] Kalt, C.: RFC 2810 Internet Relay Chat: Architecture (2000).

[9] Kalt, C.: RFC 2811 Internet Relay Chat: Channel management (2000).

[10] Kalt, C.: RFC 2812 Internet Relay Chat: Client protocol (2000).

[11] Kalt, C.: RFC 2813 Internet Relay Chat: Server protocol (2000).

[12] Kofidis, E., Regalia P. A.: On the Best rank-1 Approximation of Higher-Order Supersymmetric Tensors. SIAM J. Matrix Anal. Appl. Vol. 22, No. 3, (2002), 863–884.

[13] Kolda, T. G.: Orthogonal Tensor Decompositions SIAM Journal on Matrix Analysis and Applications. Vol. 23, No. 1, (2001), 243–255.

[14] Kolda, T. G.: A Counter Example to the Possibility of an Extension of the Eckart-Young Low-rank Approximation Theorem for the Orthogonal Rank Tensor Decomposition. SIAM Journal on Matrix Analysis and Applications Vol. 24, No. 3, (2003), 762–767.

[15] Krebs, V.: An introduction to social network analysis. www.orgnet.com/sna.html (2004) (accessed 10 February 2004).

[16] Kroonenberg, P.: Three-mode Principal Component Analysis: Theory and Applications. DSWO press. Leiden. (1983). http://www.fsw.leidenuniv.nl/ kroonenb/ (Accessed 1 June 2004).

[17] Leibovici, D., Sabatier, R.: A Singular Value Decomposition of a k-ways array for a Principal Component Analysis of multi-way data, the PTA-k. In: Linear Algebra and its Applications, (1998), 269:307–329.

[18] De Lathauwer, L., De Moor B., Vandewalle J.: On the best rank-1 and rank-(R1,R2,...,Rn) approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. Vol. 21, No. 4, (2000), 1324–1342.

[19] Mutton, P., Golbeck, J.: Visualization of semantic metadata and ontologies. In: Seventh International Conference on Information Visualization (IV03), IEEE (2003) 300–305.

[20] Mutton, P.: Piespy social network bot. http://www.jibble.org/piespy/ (2001) (accessed 14 October 2003).

[21] Natta D. V., Butler D. How Tiny Swiss Cellphone Chips Helped Track Global Terror Web. New York Times, FOREIGN DESK, Section A , Page 1 , Column 1, March 4, 2004, Thursday.

[22] Perl: http://www.perl.com. (accessed 6 June 2004).

[23] Ruskey F.: A Survey of Venn Diagrams. Electronic Journal of Combinatorics Dynamic Survey, Last updated 2001.

[24] Golub, G.H., Loan, C.F.V.: Matrix Computations. 3rd edn. The Johns Hopkins University Press, Baltimore, MD (1996).

[25] Viegas, F.B., Donath, J.S.: Chat circles. In: CHI 1999, ACM SIGCHI (1999) 9–16.

[26] Morrison A., Punin, J. and Krishnamoorthy M.S.: Real-Time Monitoring and Alerting of Web Usage. WebKDD 2004 June, 2004 (submitted).

[27] Zhang, T. and Golub, G. H.: Rank-one Approximation to Higher Order Tensors. In: Siam J. Matrix Anal. Appl. Vol.23, No. 2, (2001) 534–550.

SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION	FOR NSF USE ONLY		
Rensselaer Polytechnic Institute	PROPOSAL NO.	DURATION (months) Proposed	Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR	AWARD NO.		
Bulent Yener			

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	S	Funds Requested By	Funds granted by NSF (if different)
1. Bulent Yener - Co-Principal Investigator					$
2. Mukkai S Krishnamoorthy - Co-Principal Investigator					
3.					
4.					
5.					
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)					
7. TOTAL SENIOR PERSONNEL (1 - 6)					
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (0) POST DOCTORAL ASSOCIATES					
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)					
3. GRADUATE STUDENTS					
4. (0) UNDERGRADUATE STUDENTS					
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)					
6. (0) OTHER					
TOTAL SALARIES AND WAGES (A + B)					
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)				5,000	
2. FOREIGN				0	
F. PARTICIPANT SUPPORT COSTS					
1. STIPENDS $ 0					
2. TRAVEL 0					
3. SUBSISTENCE 0					
4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				0	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				0	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				6,000	
5. SUBAWARDS				0	
6. OTHER				38,388	
TOTAL OTHER DIRECT COSTS				44,388	
H. TOTAL DIRECT COSTS (A THROUGH G)				117,821	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)					
TOTAL INDIRECT COSTS (F&A)				39,852	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				157,673	
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				$ 157,673	$

M. COST SHARING PROPOSED LEVEL $ Not Shown AGREED LEVEL IF DIFFERENT $

PI/PD NAME	FOR NSF USE ONLY INDIRECT COST RATE VERIFICATION		
Bulent Yener	Date Checked	Date Of Rate Sheet	Initials - ORG
ORG. REP. NAME* Richard scammell			

1 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

Cumulative

ORGANIZATION	FOR NSF USE ONLY		
Rensselaer Polytechnic Institute	PROPOSAL NO.	DURATION (months) Proposed	Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR	AWARD NO.		
Bulent Yener			

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL / ACAD	Funds Requested By	Funds granted by NSF (if different)
1. Bulent Yener - Co-Principal Investigator			$
2. Mukkai S Krishnamoorthy - Co-Principal Investigator			
3.			
4.			
5.			
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)			
7. TOTAL SENIOR PERSONNEL (1 - 6)			
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)			
1. (0) POST DOCTORAL ASSOCIATES			
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)			
3. GRADUATE STUDENTS			
4. (0) UNDERGRADUATE STUDENTS			
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)			
6. (0) OTHER			
TOTAL SALARIES AND WAGES (A + B)			
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)			
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)			
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)			
TOTAL EQUIPMENT		0	
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)		5,000	
2. FOREIGN		0	
F. PARTICIPANT SUPPORT COSTS			
1. STIPENDS $ 0			
2. TRAVEL 0			
3. SUBSISTENCE 0			
4. OTHER 0			
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS		0	
G. OTHER DIRECT COSTS			
1. MATERIALS AND SUPPLIES		0	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION		0	
3. CONSULTANT SERVICES		0	
4. COMPUTER SERVICES		6,000	
5. SUBAWARDS		0	
6. OTHER		38,388	
TOTAL OTHER DIRECT COSTS		44,388	
H. TOTAL DIRECT COSTS (A THROUGH G)		117,821	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)			
TOTAL INDIRECT COSTS (F&A)		39,852	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)		157,673	
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)		0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)		$ 157,673	$

M. COST SHARING PROPOSED LEVEL $ Not Shown | AGREED LEVEL IF DIFFERENT $

PI/PD NAME	FOR NSF USE ONLY		
Bulent Yener	INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME*	Date Checked	Date Of Rate Sheet	Initials - ORG
Richard scammell			

C *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

66

Budget Justification Page

b/p

The budget includes support for 2 research assistants (one to work on visualization, one to work on multidimensional SVD). The PIs Dr Yener and Dr. Krishnamoorthy requested support each. A domestic travel budget item of $5,000 is also requested to attend the meeting and conferences. Since the research will use the departmental computing facilities we ask for $6,000 for computing services.

Current and Pending Support
For Bulent Yener

Contracts and Grants System

Current Support

Fund Number: **A11432**
Sponsor: National Science Foundation
Sponsor Reference No.: IIS-0324947
Title: ITR: Study Of Dynamically Evolving Social Groups In Communication Networks

Budget Amount: $250,000
Budget Period: 9/15/2003 - 8/31/2004
Contract Amount: $250,000
Contract Period: 9/15/2003 - 8/31/2004
Effort: Co-Principal Investigator
Academic Support: *No Charge*
Summer Support: *1 wk @ 100%*
Calendar Support: *No Charge*
Associates:

Mark K. Goldberg	Co-Principal Investigator
Mukkai S. Krishnamoorthy	Co-Principal Investigator
Malik Magdon-Ismail	Co-Principal Investigator
William A. Wallace	Co-Principal Investigator

Fund Number: **B10449**
Sponsor: National Science Foundation
Sponsor Reference No.: ANI-0340877
Title: Workshop for Pervasive Computing and Networking

Budget Amount: $55,211
Budget Period: 9/15/2003 - 8/31/2004
Contract Amount: $55,211
Contract Period: 9/15/2003 - 8/31/2004
Effort: Principal Investigator
Academic Support: *No Charge*
Summer Support: *No Charge*
Calendar Support: *No Charge*
Associates:

Pending Proposals

Proposal Number:
Sponsor:
Type:
Title:

bp

Current and Pending Support
For Bulent Yener

Contracts and Grants System

bb

Funds Requested:
Contract Period:
Effort:
Academic Support:
Summer Support:
Calendar Support:
Associates:

Current and Pending Support
For Bulent Yener

Contracts and Grants System

Petros Drineas Co-Investigator

Proposal Number: **429-04-260D**
Sponsor: Health And Human Services - National Institutes Of Health
Type: New
Title: A Novel Analytical Tool for Cancer Diagnosis
Funds Requested: $339,703
Contract Period: 8/1/2004 - 7/31/2006
Effort: Co-Principal Investigator
Academic Support: *No Charge*
Summer Support: *6 wks @ 100%*
Calendar Support: *No Charge*
Associates: Charles V. Stewart Co-Principal Investigator

Proposal Number: **493-04-265A**
Sponsor: National Science Foundation
Type: New
Title: ITR: Collaborative Research: Distributed Coordination for Joint Routing and Power Control in Multi-hop Wireless Networks
Funds Requested: $254,335
Contract Period: 9/1/2004 - 8/31/2007
Effort: Co-Investigator
Academic Support: *No Charge*
Summer Support: *4 wks @ 100%*
Calendar Support: *No Charge*
Associates:

Proposal Number: **525-04-265A**
Sponsor: National Science Foundation
Type: New
Title: Collaborative Research: Assisted Secure Multi-agent Computing with Universal Auditability
Funds Requested: $290,386
Contract Period: 9/1/2004 - 8/31/2007
Effort: Co-Investigator
Academic Support: *No Charge*
Summer Support: *6 wks @ 100%*
Calendar Support: *No Charge*
Associates:

Current and Pending Support
For Bulent Yener

Contracts and Grants System

Proposal Number:	**601-04-260A**	
Sponsor:	National Science Foundation	
Type:	New	
Title:	NeTS-NOSS: Medium Access Control Protocols for Asynchronous Sensor Networks	
Funds Requested:	$559,572	
Contract Period:	11/1/2004 - 10/31/2007	
Effort:	Co-Principal Investigator	
Academic Support:	*9 mos@)%*	
Summer Support:	*4 wks@100%*	
Calendar Support:	*No Charge*	
Associates:	Costas Busch	Co-Principal Investigator
	Malik Magdon-Ismail	Co-Principal Investigator

Current and Pending Support

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.

Investigator: **Bulent Yener**	Other agencies (including NSF) to which this proposal has been/will be submitted.

Support: ☐ Current ☒ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title: **Surveillance, Analysis and Modeling of Chatroom Communities**

Source of Support: **NSF**
Total Award Amount: $ **157,673** Total Award Period Covered: **01/01/05 - 12/31/05**
Location of Project: **RPI**
Person-Months Per Year Committed to the Project. Cal: **0.00** Acad: **0.00** Sumr: **1.00**

Support: ☐ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title:

Source of Support:
Total Award Amount: $ Total Award Period Covered:
Location of Project:
Person-Months Per Year Committed to the Project. Cal: Acad: Sumr:

Support: ☐ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title:

Source of Support:
Total Award Amount: $ Total Award Period Covered:
Location of Project:
Person-Months Per Year Committed to the Project. Cal: Acad: Sumr:

Support: ☐ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title:

Source of Support:
Total Award Amount: $ Total Award Period Covered:
Location of Project:
Person-Months Per Year Committed to the Project. Cal: Acad: Sumr:

Support: ☐ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title:

Source of Support:
Total Award Amount: $ Total Award Period Covered:
Location of Project:
Person-Months Per Year Committed to the Project. Cal: Acad: Summ:

*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.

 USE ADDITIONAL SHEETS AS NECESSARY

Current and Pending Support
For Mukkai S. Krishnamoorthy

Contracts and Grants System

Current Support

Fund Number: **A11432**
Sponsor: National Science Foundation
Sponsor Reference No.: IIS-0324947
Title: ITR: Study Of Dynamically Evolving Social Groups In Communication Networks

Budget Amount: $250,000
Budget Period: 9/15/2003 - 8/31/2004
Contract Amount: $250,000
Contract Period: 9/15/2003 - 8/31/2004
Effort: Co-Principal Investigator
Academic Support: *No Charge*
Summer Support: *1 wks @ 100%*
Calendar Support: *No Charge*
Associates:

Mark K. Goldberg	Co-Principal Investigator
Bulent Yener	Co-Principal Investigator
Malik Magdon-Ismail	Co-Principal Investigator
William A. Wallace	Co-Principal Investigator

Pending Proposals

Proposal Number:
Sponsor:
Type:
Title:

bp

Funds Requested:
Contract Period:
Effort:
Academic Support:
Summer Support:
Calendar Support:
Associates:

Current and Pending Support
For Mukkai S. Krishnamoorthy

Contracts and Grants System

Proposal Number: **382-04-506A**
Sponsor: National Science Foundation
Type: New
Title: MRI: Acquisition of Infrastrucutre For Research in Grid Computing and Multiscale Systems Computation
Funds Requested: $698,795
Contract Period: 9/1/2004 - 8/31/2007
Effort: Co-Investigator
Academic Support: *No Charge*
Summer Support: *No Charge*
Calendar Support: *No Charge*

Associates:

Mark S. Shephard	Co-Principal Investigator
Christopher D. Carothers	Co-Principal Investigator
Shekhar S. Garde	Co-Principal Investigator
Jeffrey C. Trinkle	Co-Principal Investigator
Carlos Varela	Co-Principal Investigator
Sibel Adali	Co-Investigator
Kurt S. Anderson	Co-Investigator
Georges Belfort	Co-Investigator
Kristin P. Bennett	Co-Investigator
Christopher Bystroff	Co-Investigator
Suvranu De	Co-Investigator
Jonathan S. Dordick	Co-Investigator
Jacob Fish	Co-Investigator
Joseph E. Flaherty	Co-Investigator
Mark K. Goldberg	Co-Investigator
Kenneth Jansen	Co-Investigator
Pawel Keblinski	Co-Investigator
Sanat K. Kumar	Co-Investigator
Franklin T. Luk	Co-Investigator
Antoinette Maniatty	Co-Investigator
Joyce R. McLaughlin	Co-Investigator
Don L. Millard	Co-Investigator
David R. Musser	Co-Investigator
Saroj Nayak	Co-Investigator
Rahmi Ozisik	Co-Investigator
Jong-Shi Pang	Co-Investigator
Catalin Picu	Co-Investigator
Robert L. Spilker	Co-Investigator
Charles V. Stewart	Co-Investigator
Boleslaw K. Szymanski	Co-Investigator
William A. Wallace	Co-Investigator

Current and Pending Support
For Mukkai S. Krishnamoorthy

Contracts and Grants System

	John T. Wen	Co-Investigator
	George Xu	Co-Investigator
	Bulent Yener	Co-Investigator
	Mohammed J. Zaki	Co-Investigator
	Petros Drineas	Co-Investigator
Proposal Number:	**571-04-501A**	
Sponsor:	National Science Foundation	
Type:	New	
Title:	Applying Fractal Measures of Conversation Trees to a Comparison of Open and Closed Information Cultures	
Funds Requested:	$648,788	
Contract Period:	9/1/2004 - 8/31/2007	
Effort:	Co-Investigator	
Academic Support:	*No Charge*	
Summer Support:	*2 wks @ 100%*	
Calendar Support:	*No Charge*	
Associates:	Ron Eglash	Co-Investigator
Proposal Number:	**606-04-054A**	
Sponsor:	National Science Foundation	
Type:	New	
Title:	NeTs-NOSS:Scalable Adaptive Data Gathering Using Wireless Sensor Networks	
Funds Requested:	$434,781	
Contract Period:	9/1/2004 - 8/31/2007	
Effort:	Co-Principal Investigator	
Academic Support:	*No Charge*	
Summer Support:	*4 wks@100%*	
Calendar Support:	*No Charge*	
Associates:	Koushik Kar	Co-Principal Investigator

Current and Pending Support

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.

Investigator: **Mukkai Krishnamoorthy**

Other agencies (including NSF) to which this proposal has been/will be submitted.

Support: ☐ Current ☒ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support

Project/Proposal Title: **Surveillance, Analysis and Modeling of Chatroom Communities**

Source of Support: **NSF**

Total Award Amount: $ **157,673** Total Award Period Covered: **01/01/05 - 12/31/05**

Location of Project: **RPI**

Person-Months Per Year Committed to the Project. Cal: **0.00** Acad: **0.00** Sumr: **1.00**

Support: ☐ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support

Project/Proposal Title:

Source of Support:

Total Award Amount: $ Total Award Period Covered:

Location of Project:

Person-Months Per Year Committed to the Project. Cal: Acad: Sumr:

Support: ☐ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support

Project/Proposal Title:

Source of Support:

Total Award Amount: $ Total Award Period Covered:

Location of Project:

Person-Months Per Year Committed to the Project. Cal: Acad: Sumr:

Support: ☐ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support

Project/Proposal Title:

Source of Support:

Total Award Amount: $ Total Award Period Covered:

Location of Project:

Person-Months Per Year Committed to the Project. Cal: Acad: Sumr:

Support: ☐ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support

Project/Proposal Title:

Source of Support:

Total Award Amount: $ Total Award Period Covered:

Location of Project:

Person-Months Per Year Committed to the Project. Cal: Acad: Summ:

---

*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.

 USE ADDITIONAL SHEETS AS NECESSARY

Facilities and Equipment

The RPI component of the proposed research will be conducted at the Computer Science Department Laboratories, The Center for Pervasive Computing and Networking which the PI is a Co-Director, and the Scientific Computation Research Center (SCOREC. Use of both DOE and Pittsburgh Supercomputing Center (PSC) resources will be leveraged as well. Through SCOREC, we currently have access to PSC resources. In addition to PSC resources, SCOREC also has remote access to the computational facilities at the San Diego Super Computer Center.

The Computer Science Department operates its own computer facilities which consists largely of 120 Sun workstations. These systems are the department's general purpose computing facilities which are used primarily by the faculty and graduate students. Additionally, department's researchers have very open access to a wide variety of equipment, including a large distributed-memory parallel computer, a large symmetric multiprocessor machine (SGI Origin 2000) and many general-purpose workstations.

Additionally, the department just received 8, quad Itanium systems with 16 GB of RAM each via an NSF equipment grant. We plan to make use of this state-of-the-art hardware in the development of our system before scaling it up to execute on the DOE and PSC supercluster platforms.

Our newly formed Center for Pervasive Computing and Networking research owns a distributed computing and networking laboratory equipped with 40 Netfinity IBM machines with 700 MHz processors (donated recently by IBM Corp.) and network routers, switches and hubs (donated by Cisco System Inc.). In addition, this laboratory includes several Sun Ultra 10's purchased from research funds for Research Assistants.

In addition to the Itanium cluster, this project make use of high-performance Linux cluster. This cluster has a mix of computing hardware starting with 8, dual processor PC servers interconnected with a high-speed Myrinet network. Each Myrinet link provides 1.6 gigabit/second of bandwidth. Each PC server contains two, 400 MHz Pentium II processors, with 256 MB of memory and 27 GB of disk space. In addition, the Scientific Computation Research Center (SCOREC) added four, quad, 500 MHz Pentium III processor systems to the Myrinet cluster, and finally 3, dual hyperthreaded Xeon Pentium IV systems with 6 GB of RAM each with a head-end fileserver with 1.7 TB of RAID 5 disk space, bring the total number of processors to 40.