George Mason University
Office of Sponsored Programs
4400 University Drive MS 4C6
Fairfax, VA 22030
(703) 993-2988
Fax # (703) 993-2296

2003 JUL 31 AM 10:00

July 29, 2003

Director
U.S. Army Research Office
ATTN: AMSRL-RO-RI (DAAD19-00-R-0010)
4300 South Miami Boulevard
Durham, NC 27703-9142

RE: DAAD19-00-R-0010

Dear Sir or Madam:

Attached please find the signed Forms 51, 52A, 99, and certifications for a proposal entitled, **"Analytic and Graphical Methods for Streaming Data with Applications to Netcentric Warfare"** that was submitted electronically on behalf of Dr. Edward Wegman, Center for Computational Statistics (CCS), George Mason University.

If you have any questions regarding the technical content of this proposal, please feel free to contact Dr. Wegman at 703/993-1691 or ewegman@gmu.edu . Questions regarding budget or university policy should be directed to Lauren Magruder, Grants Administrator, Office of Sponsored Programs, at 703/993-2978 or lmagrude@gmu.edu .

Sincerely,

Ann T. McGuigan

Ann T. McGuigan, Ph.D.
Director, Office of Sponsored Programs

Enclosures
Cc: R. Launer
E. Wegman

U.S. Postal Service:
George Mason University
Office of Sponsored Programs, MSN 4C6
4400 University Drive
Fairfax, VA 22030

Federal Express Delivery:
GMU Office of Sponsored Programs
10517-C Braddock Road
Suite 2501
Fairfax, VA 22032

Solicitation Number:	PROPOSAL COVER PAGE	45810

1. SUBMIT TO:
Director
U.S. Army Research Office
ATTN: AMSRL-RO-RI
P.O. Box 12211
Research Triangle Park, NC 27709-2211

2003 JUL 31 A 8:08

2. For consideration by:
- ☐ Biology/Life Sci
- ☐ Chemistry
- ☐ Computer Science
- ☐ Electronics
- ☐ Mechanical
- ☐ Environmental Sciences
- ☐ Sensors & Electron Dev
- ☐ Materials
- ☑ Mathematics
- ☐ Physics
- ☐ Comp & Info Sci
- ☐ Wepons & Mtls Sci
- ☐ Human Rsch & Eng
- ☐ Surv/Lethality

3. Is this proposal being submitted to another Federal Agency?
☐ No ☑ Yes If Yes, list the agency:
Office of Naval Research

4. Is applicant delinquent on any Federal Debt?
☐ Yes (Attach explanation) ☑ No

5. Proposal Valid Until (min of 6 mos):
6 Months from date of submission

6. Entity Identification Number (EIN) or Taxpayer Identification Number (TIN)
54-0836354

7. Data Universal Numbering System (DUNS No.):
07-781-7450

8. Commercial and Government Entity (CAGE) Code: 7X764

9. Name of organization to which award should be made:
George Mason University
Office of Sponsored Programs
4400 University Drive, MS 4C6
Fairfax, VA 22030

10. Administrative Address of Organization (if different):
N/A

11. Branch/Campus/Other Component (where work is performed, if different): N/A

12. Submitting Organization's Contract/Grant Administration Office:
Office of Sponsored Programs
4400 University Drive, MS 4C6, Fairfax, VA 22030

13. Submitting Organization's Audit Office:
Office of Naval Research

2003 JUL 31 A 8:07

14. Submitting Organization: (Check all that apply)
- ☐ For Profit: ☐ Large ☐ Small ☐ Disadvantaged ☐ 8a ☐ Women-Owned ☐ Foreign ☐ Individual
- ☑ Educational: ☐ HBCU ☐ Minority Institution ☐ Hispanic ☐ Indian Tribal ☑ State ☐ Private ☐ Foreign ☐ FDP
- ☐ Hospital: ☑ Public ☐ Private ☐ Nonprofit ☐ For Profit
- ☐ Nonprofit
- ☐ Not-For-Profit
- ☐ Other (Specify)

15. Check appropriate box(es) if this proposal includes any of the items listed below:
- ☐ Human Subjects
- ☐ Recombinant DNA
- ☐ Vertebrate Animals
- ☐ Genetically Engineered Organisms
- ☐ National Environment Policy Act
- ☐ Limited Rights Data
- ☐ Disclosure of Lobbying Activities
- ☐ Unlimited Rights
- ☐ Historical Places
- ☐ Govt Purpose Rights Data
- ☐ GFE ☐ GFD
- ☐ Proprietary Data
- ☐ GFI ☐ GFP
- ☐ Ozone Depleting Substances

16. Proposed Amount: $273,941

17. Proposed Duration (1-60 mos): 36 months

18. Proposed Start Date: October 16, 2003

19. Type of Award Proposed:
- ☑ Single Investigator
- ☐ Young Investigator Program
- ☐ Short Term Innovation Rsch
- ☐ Research Instrumentation
- ☐ Conference/Symposia
- ☐ Other (Specify):

20. Title of Proposed Project: Analytic and Graphical Methods for Streaming Data with Applications to Netcentric Warfare

21. Principal Investigator (PI)/Project Director (PD) Department and Postal Address:
Dr. Edward Wegman, Center for Computational Statistics, 4400 University Drive, MS 4A7, Fairfax, VA 22030

TYPED NAMES	TELEPHONE NUMBER	FACSIMILE NUMBER	ELECTRONIC MAIL ADDRESS
22. PI/PD Dr. Edward Wegman	703-993-1691	703-993-1700	ewegman@gmu.edu
23. CO-PI/PD			
24 a. Primary Administrative Representative Authorized to Conduct Negotiations: Ann T. McGuigan, Ph.D.	703-993-2988	703-993-2296	amcguiga@gmu.edu
24 b. Alternate Administrative Representative Authorized to Conduct Negotiations: Mindy Barnhart	703-993-4573	703-993-2296	mbarnhar@gmu.edu

25 a. Authorized Representative Signing for Applicant Organization:
Ann T. McGuigan, Ph.D.
25 b. Title: Director, Office of Sponsored Programs

25 c. By signing and submitting this proposal, the offeror is providing the certifications contained in this BAA.
25 d. Signature Ann T. McGuigan Date: 7-29-03

PROTECTION OF PROPRIETARY INFORMATION DURING EVALUATION AND AFTER AWARD

It is the policy of the US Army Research Office to treat all research proposals as privileged information prior to award and to disclose the proposal contents only for evaluation purposes. Technical evaluation of these proposals normally are made by highly qualified scientists from the Government and leading scientists and other preeminent experts outside the Government to ascertain their merits. If you wish, you may restrict the evaluation of your proposal to only scientists within the Government. To do so may prevent it from receiving an evaluation by those most qualified to evaluate it. Therefore, we ask permission to send your proposal outside the Government, if necessary, to obtain an unrestricted evaluation.

All reviewers are made aware that proposals sent to them for evaluation shall not be duplicated, used, or disclosed in whole or in part for any other purpose without the written permission of the offeror.

You should be aware that despite all precautions, we can protect the confidentiality of proprietary information contained in proposals only to the extent that it is exempt from disclosure under the Freedom of Information Act (FOIA; 5 U.S.C. 552). Generally, Exemption 4 of the FOIA [5 U.S.C. 552(b)(4)] will protect from release information submitted to the Government that constitutes either (1) a trade secret or (2) commercial or financial information which is privileged or confidential. Any such proprietary information contained in your proposal should be marked in accordance with FAR 15.608.

Please complete the following statement indicating your proposal treatment preference during the evaluation phase and, should a contract be awarded, afterwards.

STATEMENT OF DISCLOSURE PREFERENCE

George Mason University (Organization) in submitting proposal titled Analytic and Graphical Methods for Streaming Data with Applications to Netcentric Warfare with Edward J. Wegman (Name)

as Principal Investigator, requires the following procedure be used during its evaluation:

(X) Permission is hereby granted to the US Army Research Office to evaluate this proposal in accordance with its normal procedures which may include evaluation by reviewers both within and outside the Government.

( ) Restrict the evaluation of the above proposal to Government reviewers only.

To reduce administrative requirements, you may complete the following information by checking the appropriate blocks:

If this proposal results in a contract or grant, the offeror grants ARO the authority to release the following portion of its proposal in response to requests under the Freedom of Information Act, only after such award, without prior contact with the offeror.

Cover Page	☐ Yes	☐ No	Statement of Work	☐ Yes	☐ No
Project Abstract	☐ Yes	☐ No	Biographical Sketch	☐ Yes	☐ No
Technical Proposal	☐ Yes	☐ No	Cost Proposal/Budget	☐ Yes	☐ No
Current & Pending Support	☐ Yes	☐ No	Bibliography	☐ Yes	☐ No

7-29-03
Date

Ann I. McGuigan
Signature of person authorized to sign for organization

Date

Edward J. Wegman
Signature of Principal Investigator

ARO Form 52A
(Revised Oct 98)
USE THIS FORM FOR EDUCATIONAL INSTITUTIONS AND NON-PROFIT ORGANIZATIONS

PROPOSAL

TABLE OF CONTENTS

PROTECTION OF PROPRIETARY INFORMATION DURING EVALUATION AND AFTER AWARD

It is the policy of the US Army Research Office to treat all research proposals as privileged information prior to award and to disclose the proposal contents only for evaluation purposes. Technical evaluation of these proposals normally are made by highly qualified scientists from the Government and leading scientists and other preeminent experts outside the Government to ascertain their merits. If you wish, you may restrict the evaluation of your proposal to only scientists within the Government. To do so may prevent it from receiving an evaluation by those most qualified to evaluate it. Therefore, we ask permission to send your proposal outside the Government, if necessary, to obtain an unrestricted evaluation.

All reviewers are made aware that proposals sent to them for evaluation shall not be duplicated, used, or disclosed in whole or in part for any other purpose without the written permission of the offeror.

You should be aware that despite all precautions, we can protect the confidentiality of proprietary information contained in proposals only to the extent that it is exempt from disclosure under the Freedom of Information Act (FOIA; 5 U.S.C. 552). Generally, Exemption 4 of the FOIA [5 U.S.C. 552(b)(4)] will protect from release information submitted to the Government that constitutes either (1) a trade secret or (2) commercial or financial information which is privileged or confidential. Any such proprietary information contained in your proposal should be marked in accordance with FAR 15.608.

Please complete the following statement indicating your proposal treatment preference during the evaluation phase and, should a contract be awarded, afterwards.

STATEMENT OF DISCLOSURE PREFERENCE

George Mason University (Organization) in submitting proposal titled Analytic and Graphical Methods for Streaming Data with Applications to Netcentric Warfare with Edward J. Wegman (Name)

as Principal Investigator, requires the following procedure be used during its evaluation:

(X) Permission is hereby granted to the US Army Research Office to evaluate this proposal in accordance with its normal procedures which may include evaluation by reviewers both within and outside the Government.

( ) Restrict the evaluation of the above proposal to Government reviewers only.

To reduce administrative requirements, you may complete the following information by checking the appropriate blocks:

If this proposal results in a contract or grant, the offeror grants ARO the authority to release the following portion of its proposal in response to requests under the Freedom of Information Act, only after such award, without prior contact with the offeror.

Cover Page	☐ Yes	☐ No	Statement of Work	☐ Yes	☐ No
Project Abstract	☐ Yes	☐ No	Biographical Sketch	☐ Yes	☐ No
Technical Proposal	☐ Yes	☐ No	Cost Proposal/Budget	☐ Yes	☐ No
Current & Pending Support	☐ Yes	☐ No	Bibliography	☐ Yes	☐ No

Date — Signature of person authorized to sign for organization

Date — Signature of Principal Investigator

ARO Form 52A
(Revised Oct 98)
USE THIS FORM FOR EDUCATIONAL INSTITUTIONS AND NON-PROFIT ORGANIZATIONS

PROJECT ABSTRACT

The Project Abstract shall include a statement of objectives, methods to be employed, and the significance of the proposed activity to the advancement of knowledge or education. Avoid use of the first person to complete this summary. DO NOT EXCEED ONE PAGE. The abstract should be suitable for release under the Freedom of Information Act, 5 U.S.C. 552, as amended.

This proposal focuses on a new data structure, namely massive streaming data. We propose to develop recursive algorithms and evolutionary graphics for handling massive streaming data. Streaming data is essentially a new data acquisition paradigm, in which data becomes constantly available. Older data has less value and therefore must be discounted. Strategies for discounting are proposed as well as strategies for multi-scale resolution of data streams. The particular example we have in mind is streaming Internet packet headers, although theoretical and practical results will not be limited to this form of data. However, network traffic data are especially important to military and the U. S. Army in particular as Netcentric Warfare and joint operations between services and with allies become increasingly important. The ultimate goal of these techniques is to detect intrusion and fraud in streaming data systems.

Analytic and Graphical Methods for Streaming Data with Applications to Netcentric Warfare

1. Introduction

Netcentric warfare concepts evolved from the 1991 Gulf War experience and have been defined by several sources. The Committee on Network-Centric Naval Forces of the Naval Studies Board of the National Research Council (2000) defined Network Centric Operations as follows:

> "Network-centric operations (NCO) [are] military operations that exploit state-of-the-art information and networking technology to integrate widely dispersed human decision makers, situational and targeting sensors, and forces and weapons into a highly adaptive, comprehensive system to achieve unprecedented mission effectiveness."

They go on to observe that:

> "In one way or another all military operations will be joint. That is, systems and forces from all the services and from National agencies will contribute to the U. S. Armed Forces' operations in ways that vary with the circumstances."

The reliance for total mission effectiveness on information and networking technology brings concomitant vulnerabilities. Modern information and networking technology could be destroyed relatively easily by an electro-magnetic pulse. More likely however, is the interception of networking technology by clever hackers, even on secure communication networks. Much of the communication technology especially from sensor platforms is radio based and subject to spoofing and other methods of distorting the overall situational awareness.

Data mining, when compared with the classical statistical analysis paradigm, shows a substantial change of perspective. Instead of relatively small, low-dimensional, homogeneous data sets derived from a carefully designed sampling procedure, awareness of the issues in data mining has led many researchers to consider massive, high-dimensional datasets that may not satisfy traditional homogeneity assumptions. In addition, data sets used in a data mining context often have been collected for administrative or other purposes having little to do with the purposes for which data mining techniques are applied. Nonetheless, even among those aware of the issues of computational complexity and massive data sets, the typical mindset is that we have a dataset of fixed size n, however large n might be. However, we believe there is a new

data collection paradigm looming on the horizon, to wit, data is streaming, coming available continuously, and realistically not all of it can be stored. It is clear that the netcentric operations idea will generate a streaming of the type suggested by what we identify as the new paradigm. Moreover as new data becomes available the value of older data diminishes. Almost a given with streaming data is that data are not time homogeneous. Indeed this is a strong contrast with the conventional perspective on homogeneously sampled data.

In addition to data generated by networks of computers, examples of this class of data abound. Point of purchase data, telephone traffic data, and credit card use data are all examples. The data on which we will focus in this proposal is Internet traffic data. Techniques for the use and analysis of such data must of necessity be somewhat different from fixed sample size data. Because the data cannot be stored conveniently, recursive algorithms that update the desired statistic using an incoming piece of data and then discard that piece of data are appropriate. Data visualization methods have more recently emphasized dynamic graphics, i.e. animation of the graphic using rotation, grand tours, mapping of variables into time, and so on, but always with an eye on a fixed dataeset size. We propose what we like to call *evolutionary graphics*, i.e. graphics which evolve as a function of new data being added. The combination of recursive algorithms and evolutionary graphics will provided a fundamental approach for analyzing streaming data.

As a prototype for developing tools for streaming data, we have launched on a data collection effort with the agreement of the George Mason University's CIO to capture all header information for all Internet traffic in and out of the University. This comprises primarily TCP, UDP and ICMP packets. We have installed sniffer and analysis machines and are capable of locally recording up to a terabyte of traffic data. Preliminary experiments within our small statistics subnet indicate traffic of 65,000 to 150,000 packets per hour. We are currently collecting about 1.2 to 3.0 gigabytes of header information per hour in the larger University context. Ultimately, we are interested in real-time detection of intrusion attacks so that analysis methods for steaming data are necessary. In the next sections, we will describe some background on TCP/IP traffic, indicate some recursive methods capable of handling streaming data and give some suggestions for analytic algorithms and visualization procedures we hope to develop under this proposal.

2. Scoping the Problem, IP Header Information

2.1 IP Addressing

Most of us have seen some versions of the Internet Protocol (IP) addresses or more precisely, the IP version 4 (IPv4) addresses. An IPv4 address is a 32-bit number usually represented as 4 dotted fields, i.e. *field1.field2.field3.field4*. Each field is essentially an 8-bit byte and for this reason is often called an octet. In principle, these IP addresses uniquely identify a machine. This in fact is not true strictly speaking. In wireless networks and other settings, machines may have dynamically assigned IP addresses, which remain fixed for the duration of a

session on a machine (as long as the machine is turned on and the machine is in the network), but which address may be reassigned to another machine when the session is ended. In theory there are $2^{32} = 4,294,967,296$ addressable machines.

The configuration of the IPv4 address identifies the type of network. In a Class A network, field 1 identifies the network and fields 2-4 identify a specific host. Class A networks can be identified by having field 1 smaller than 128, e.g. 1.1.1.1. The most typical type of network is a Class B network in which field1.field2 identifies the network and field3.field4 identifies the specific host. Often field3 identifies a subnet. Field1 must be at least as large as 128, e.g. 130.103.40.210. In a Class C network, field1.field2.field3 identifies the network and field4 identifies the host., e.g. 192.9.200.15. In a Class C network, only 256 unique machines are addressable. One can immediately see issues that arise in graphical displays. Even high resolution graphical displays may display 1600 pixels across whereas only two fields (two octets) amount to 65,536 discrete entries. Thus displaying only two fields in a graphic would already create severe resolution problems without some sort of zoom and pan capability.

One might think that 4 billion Internet addresses would be sufficient. However, since many are reserved and not all are actually used, a new standard, IP version 6 (IPv6) is being introduced. An IPv6 address is a 128-bit address arranged as 8 groups of 16 bit numbers separated by colons, e.g. EFDC:BA62:7654:3201:AFDC:BA72:7654:3210. Leading zeros may be omitted so that 1060:0000:0000:0000:0006:0600:200C:326B = 1060:0:0:0:6:600:200C:326B. Any sequence of single zeros and colons may be replaced by a double colon so that 1060:0:0:0:6:600:200C:326B = 1060::6:600:200C:326B. All IPv4 addresses fit into an IPv6 address of the form ::****:****. For example, 130.103.40.5 in IPv6 notation is ::8267:2805. Note that 130 in decimal notation is 82 in hexadecimal, 103 in decimal is 67 in hexadecimal, 40 is 28 in hexadecimal and 5 in decimal is also 5 in hexadecimal. Hybrids are allowed so ::130.103.40.5 is acceptable. If dealing with IPv4 addresses strains the data visualization, then the number of IPv6 addresses, $2^{128} = 3.4028 \times 10^{38}$ greatly compounds that problem.

2.2 Ports

Network-connected hosts have so-called ports, which may be viewed as extensions to their respective IP addresses in the sense that they identify what services and applications are communicating. Ports are a logical rather than physical connections, and many of the ports have a standard service, application, and/or vendor associated with them. There are $2^{16} = 65,536$ ports for each host, divided into three ranges: well-known ports (0-1023); registered ports (1024-49151); and dynamic and/or private ports (49152-65535). Ports are used in TCP to name the ends of logical connections, which (potentially) carry long-term conversations. The so-called well-known ports are assigned by the Internet Assigned Numbers Authority (IANA) and on most systems can only be used by system or root processes or by programs executed by privileged users. The registered ports are listed by the IANA and on most systems can be used by ordinary user processes or programs executed by ordinary users. Some well-known

standard services and associated ports are file transfer protocol (ftp) on port 21, secure shell (ssh) on port 22, telnet on port 23, mail (smtp) on port 25, web services (http) on port 80, mail (pop3) on port 110. Sun network file system (nfs) is registered for port 2049. Even DirecTV and AOL have registered ports associated with them. For more details see http://www.IANA.org. Unprotected (open) ports allow for possible intrusion or malicious use. See Marchette and Wegman (2003) for a description of some of these strategies. Scanning a host for unprotected ports is a strategy that hackers may use to break into a machine. Not all 65,000 ports are typically scanned nor are they necessarily scanned sequentially. Just as with IP addresses, the total number of ports makes visualization of all of them difficult.

2.3 Packet Structure

The basic data structure on a network is the packet. All communications, whether they be email, web, chat or a remote login, are encapsulated in a series of packets. Each packet consists of a header of routing information (basically, the to and from address of the packet) plus the data. Transactions are broken into a series of relatively small packets, and sent out on the network. It is important to realize that packets are nominally dynamically routed. Each router determines the best next hop for the packet and sends it along to the next router closer to the destination. This means that different packets of the same communication can take different routes, and thus do not necessarily arrive in the order in which they are sent.

Statistically speaking, our interest will generally be focused on what is technically referred to as the IP datagram (versus the IP packet), the unit of end-to-end transmission before fragmentation and after reassembly if required by limitations imposed by the physical network over which the encapsulated data must pass. A packet on the other hand is the unit of data that is actually passed "on the wire," which under fragmentation reflects only a portion of the application generated information being transmitted to the destination host. Observations on the size of data segments or transmission counts from or to a given IP address or port, for example, would most likely refer to datagrams being sent and received. Malformed fragments are of interest because they may represent malicious intent.

The basic protocol of the Internet is IP. This is a non-reliable protocol (meaning that there are no reliability guarantees made by the protocol). Other protocols are available to provide these guarantees, and these are implemented within the data portion of the IP packet. An IP packet contains the basic address information, source and destination IP address, a unique identifier for the packet, the protocol of the data portion of the packet, fragmentation information, and various options for routing.

The most important fields in the IP packet are the source and destination IP addresses. These are used to identify the sending and receiving hosts. The other important fields are the flags and fragment offset fields. As mentioned previously, packets that are too big may be broken up into smaller packets, called fragments. These fields are used to ensure that the packets can be reassembled at the destination. If a packet is fragmented, the first fragment

consists of a fragment offset of 0 and the bit corresponding to the "more fragments" (MF) flag is set. Subsequent fragments set the fragment offset field to the value corresponding to their place in the packet. All but the last packet have the MF flag set. By assembling the packet as indicated by the fragment offset field, the original packet is reassembled. All fragments for a given packet have the same identification field as the original, so that the receive can determine which fragments belong with which packet.

The main protocols that we are concerned with, beyond IP, are the Internet Control Message Protocol (ICMP), User Datagram Protocol (UDP) and Transmission Control Protocol (TCP). ICMP is used for error messages and diagnostics, while UDP and TCP provide the basic one- and two-way communication channels used by most applications.

The IP protocol does not provide any guarantees of delivery or reliability. A packet is sent off, and if it is received there is no necessity for the receipt to be acknowledged to the sender. If the packet is lost, it is lost, with no mechanism for resending. There are error mechanisms built in, using the ICMP protocol, in which packets that cannot be delivered result in an error sent to the sender, but beyond this, any desired reliability is implemented in the other protocols.

The vast majority of traffic on the Internet is either UDP or TCP. UDP provides a one-way connection with no guarantee of service, while TCP provides two-way connections with assurance: packets which are not delivered are resent. Both protocols implement the concept of connection ports, which can be thought of as an extra two bytes of addressing. This provides a unique communication path between the two computers.

TCP implements the two-way connection via an initiating handshake with sequencing controlled by sequence numbers. The main fields of interest are the port numbers, sequence numbers, and flags. The port numbers determine the applications associated with the connection, and allow the computers to have several sessions between the applications without mixing them up.

2.4 Sniffers

In order to study Internet traffic, especially to investigate attacks on our systems, it is desirable to capture packet headers. Generally we have no interest in the data content of the packets because packets are usually fragmented and often encrypted. Thus to reconstruct the full details of a single session is a formidable task and often not a fruitful exercise anyway. However, source and destination IP and source and destination ports and packet type can be used to glean a great amount of information about the nature of traffic and if it is malicious or benign.

In order to capture the traffic, anonymous surveillance machines are installed typically just outside the firewall and monitor traffic flowing through the point. Programs such as tcpdump capture the header information. In the case at George Mason University, our sniffer machine

captures 68 bytes of header infromation for all Internet traffic in and out of the University. To give some sense of scale, we capture between 1.2 to 3.0 gigabytes of header data per hour or approximately 26 terabytes per year of just header data. According to our studies, even our relatively small statistics network generates more than 150,000 packets per hour (during a relatively quiescent examination period).

3 Recursive Algorithms

Because the volume of internet traffic is streaming and so extensive, it is essentially impossible to store all the data, even with multi-terabyte storage capability, except for some limited amount. Of course, because the nature of the data is not generally homogeneous, it is also desirable to discount older data. In this section we discuss recursive and discounting algorithms.

3.1 Moment and Count Calculations

Moments are important, of course, because they generally characterize a distribution. Moreover as we shall see in the next subsection, they also lead us to an optimal way of compressing data. Suppose we now agree that a sequence X_i, $i = 1,2,3,\ldots$ represents an incoming stream of data. Then the traditional $\overline{X}_n$ can be computed recursively by the formula

$$\overline{X}_n = \frac{n-1}{n}\overline{X}_{n-1} + \frac{X_n}{n}.$$

Similarly, moments of all orders can be computed recursively by

$$\sum_{i=1}^{n} X_i^k = \sum_{i=1}^{n-1} X_i^k + X_n^k.$$

Of course, these are essentially trivial computations as is the count computation. Clearly the asymptotic statistical properties in a stationary setting are identical with the non-recursive formulation of the estimators. However, as sample sizes increase, round-off considerations become significant and rescaling could become an issue, i.e. for very large n, the ratio $\frac{n-1}{n}$ becomes essentially indistinguishable from 1 and division by n is limited to the maximum floating point number allowed by the computer and its floating point algorithms. Moments, and means and counts particularly, become important in the construction of a pseudo-samples by geometric quantization.

3.2. Quantization

Khumbah and Wegman (2003) have suggested a method for creating pseudo-samples by geometric quantization. Although storing all of the data is impossible, it may be possible to aggregate data at a sufficiently fine level that is storable and may be useful from statistical

perspective. The idea is simple enough, tessellate the sample space at a sufficiently fine level so that incoming observations may be binned. The binning boundaries must not depend on the data so there are basic requirements that must be met:

1. The tessellations must be space filling so that every observation may be assigned to a tile.
2. It is extremely desirable to have congruent tiles so as to simplify computation.
3. The tiles should be as spherical as possible so as to minimize distortion.

The idea is that there should be k tiles where k is some large number so that the geometric size of the tiles is quite small and so that departure of an observation from its aggregated representor is minimal. In fact, we effectively do this all the time by representing a continuous image by high resolution digital images on a screen with very tiny pixels.

A key notion discussed in Khumbah and Wegman is the calculation of the representor for a tile in the tessellation. If Q is the quantizer, then the key idea is that $E[X|Q = y_j] =$ mean of the observations in the jth tile $= y_j$. In other words, $E[X|Q]=Q$ and the quantizer is said to be *self consistent*. Many important properties follow from self consistency.

1. $E[X]=E[Q]$.
2. If $\widehat{\theta}$ is a linear unbiased estimator of θ, then so is $E[\widehat{\theta}\,|Q]$.
3. If h is a convex function, then $E[h(Q)]\leq E[h(X)]$. In particular $var(Q)\leq var(X)$.
4. $E[X-P]^2 \geq E[X-Q]^2$ for any other quantizer P.

The fact that means and counts can be computed recursively implies that the value y_j the representor of the jth tile can be computed recursively and that a pseudo-sample of size k can be built from streaming data, where the pseudo-sample size is always fixed. Provided that the tiles are sufficiently small, Khumbah and Wegman (2003) have preliminary indications that the pseudo-sample is essentially indistinguishable from the real sample based on small scale studies. However, this quantizing algorithm has not been implemented in an operational sense and many of the theoretical properties are still left unaddressed.

3.3 Density Estimation

The traditional nonparametric kernel density estimator

$$\widehat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{x-X_i}{h_n})$$

has the great disadvantage of not being recursively computable. This makes this awkward for n very large and essentially useless for streaming data. Wolverton and Wagner (1969) and apparently independently Yamato (1971) introduced a recursive form of the kernel estimator:

$$f_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i} K(\frac{x-X_i}{h_i})$$

which may be reformulated in recursive form as

$$f_n^*(x) = \frac{n-1}{n} f_{n-1}^*(x) + \frac{1}{nh_n} K(\frac{x-X_n}{h_n}).$$

This formulation makes a kernel estimator appropriate for streaming data. An alternate formulation was given in Wegman and Davies (1979)

$$f_n^\dagger(x) = (nh_n)^{-\frac{1}{2}} \sum_{i=1}^{n} (nh_i)^{-\frac{1}{2}} K(\frac{x-X_i}{h_i})$$

which has a recursive form

$$f_n^\dagger(x) = \frac{n-1}{n}(\frac{h_{n-1}}{h_n})^{\frac{1}{2}} \quad f_{n-1}^\dagger(x) + \frac{1}{nh_n} K(\frac{x-X_n}{h_n}).$$

Although the bandwidth is adjustable and depends on the order in which the observations are presented, these estimators have asymptotic strong consistency and, in fact, Wegman and Davies (1979) demonstrate the exact rates of strong convergence for both of these kernel-variant estimators. The conditions are rather tedious, but generally relatively mild. See Wegman and Davies for details. Although Wegman and Davies (and for that matter Wolverton and Wagner and Yamato) focus on the one-dimensional density estimation problem, these methods and results have not been extended to the critical multidimensional case. The simplicity of these recursive formulations make them accessible for streaming data.

Priebe (1994) also demonstrates a recursive algorithm for density estimation based on mixture models. The procedure can be initiated by taking a single pair of observations and calculating and using these to estimate the parameters of a normal density function. Suppose then $\widehat{f}(\boldsymbol{x})$ is the initial density estimate. The general idea of Priebe's procedure is to either create a new mixture term if the new observation is far away from existing mixture terms or to update an existing mixture term if the next observation is sufficiently close to one of the existing mixture terms. Suppose now that we have applied his methods and have $\widehat{f}(\boldsymbol{x})$ with N mixture terms.

$$\widehat{f}(\boldsymbol{x}) = \sum_{i=1}^{N} \pi_{i,n} \phi(\boldsymbol{x}, \theta_{i,n}).$$

Typically, ϕ is taken to be a normal density with parameters given by $\theta_{i,n} = (\boldsymbol{\mu_{i,n}}, \boldsymbol{\Sigma_{i,n}})$ and mixing weights given by the π_i. Then the update rule is as follows:

$$\tau_{i,n+1} = \frac{\pi_{i,n}\phi(\boldsymbol{x}_{n+1}, \theta_{i,n})}{\sum_{t=1}^{N} \pi_{t,n}\phi(\boldsymbol{x}_{n+1}, \theta_{i,n})}$$

$$\pi_{i,n+1} = \pi_{i,n} + \frac{1}{n}(\tau_{i,n+1} - \tau_{i,n})$$

$$\boldsymbol{\mu}_{i,n+1} = \boldsymbol{\mu}_{i,n} + \frac{\tau_{i,n+1}}{n\pi_{i,n}}(\boldsymbol{x}_{n+1} - \boldsymbol{\mu}_{i,n})$$

$$\boldsymbol{\Sigma}_{i,n+1} = \boldsymbol{\Sigma}_{i,n} + \frac{\tau_{i,n+1}}{n\pi_{i,n}}[(\boldsymbol{x}_{n+1} - \boldsymbol{\mu}_{i,n})(\boldsymbol{x}_{n+1} - \boldsymbol{\mu}_{i,n})^{\dagger}].$$

Here $\tau_{i,n+1}$ is the estimated posterior probability of x_{n+1} belonging to the ith component. The remainder of the update terms provide for parameter updates. Please note that we have used bold to indicate that these are multivariate densities. A new term may be created if the new observation is sufficiently far away from any of the existing terms. This is usually measured in terms of Mahalanobis distance. If the probability of creating a new term is one, then this is essentially a traditional kernel estimator. If the probability of adding a new term is zero, then Priebe's adaptive mixtures estimator is essentially a standard parametric normal estimator. Usually, this estimator creates many spurious terms, but usually yields a good fit. The Priebe adaptive mixtures estimator, while apparently considerably more computationally complex than the kernel estimator, is in fact a good alternative because it need not be computed on a grid as the kernel estimator must. Thus in higher dimensions, it avoids the need to compute the estimator on a large grid. The mixture density also has the advantage of interpretability for clustering. Somewhat less is known about the asymptotic properties for the adaptive mixtures estimator than for the recursive kernel estimators.

In addition to the fact that properties of the Gaussian-based adaptive mixtures have less well developed theoretical properties, unfortunately the Gaussian mixtures do not form in general an orthonormal basis. A looming theoretical need for adaptive mixtures is a version with orthonormal bases such as wavelets to reduce or eliminate spurious terms.

3.4 Discounting Old Data

The fundamental premise of streaming data is that the generating structure for the data is changing with time. Hence, traditional asymptotic results mentioned above have limited value and should be taken as performance indicators, but not as serious data mining tools. A long standing device for discounting data is the use of exponential smoothing. Consider

$$Y_t = \sum_{i=0}^{\infty} (1-\theta)\,\theta^i\, X_{t-i}^k,\ 0 < \theta < 1$$

which may be reformulated in recursive form as

$$Y_t = (1-\theta)X_t + \theta Y_{t-1}.$$

First suppose that X_t has stationary momentsso that $E[X_t^k] = E[X^k]$. Then, $E[Y_t] = \sum_{i=0}^{\infty} (1-\theta)\,\theta^i E[X_{t-i}^k] = E[X^k](1-\theta)\sum_{i=0}^{\infty}\theta^i = E[X^k]$. Thus in stationary cases the exponential smoothed Y_t has exactly the same expectation as X^k would. Notice that for θ close to 1, the exponential smoother places the heaviest weight on the historical value of Y_{t-1} and little weight on the most recent value of X_t^k. This holds for all moment calculations including the mean and may be adapted to other mathematical structures as well. On the other hand if, θ is close to zero, the exponential smoother places most weight on the recent observation X_t^k and little weight on the historical value. By carefully adjusting θ we may also adjust how quickly or how slowly the exponential smoother adapts to the new data. This will depend of course on the time-scale of the changes in our case of Internet traffic. This also suggests that we could use a spread of θ values to do a multiresolution analysis of the streaming data.

Finally we note that this strategy may be applied to the recursive kernel density estimator. Note that in the Wolverton-Wagner version of the recursive kernel estimator that $\frac{n-1}{n}$ plays the role of θ and that $\frac{1}{n}$ plays the role of $(1-\theta)$. This suggests that we could formulate a new recursive kernel density estimator with adjustable discounting of old data

$$f_n^*(x) = \theta\, f_{n-1}^*(x) + \frac{(1-\theta)}{h_n} K\left(\frac{x - X_n}{h_n}\right)$$

where of course $0 < \theta < 1$. Indeed, it may be desirable to adjust θ as a function of n. This needs additional exploration as well.

An alternate traditional scheme for discounting older data is to employ a moving window and only retain the last say m observations. This moving window may be also combined with a taper to discount older observations within the moving window.

4 Evolutionary Graphics

Just as streaming data implies that there is no predetermined sample size for recursive algorithms, so too must the nature of graphic representation of data be adjusted. We propose to use the language *evolutionary graphics* rather than dynamic graphics. The latter implies a fixed sample size that is animated by rotations, grand tours, mapping a variable into time and so on. The key element is that the data set size is fixed no matter how large. With evolutionary graphics

we mean to imply that the graphic evolves with time changing along with the streaming data. Naturally it is difficult to illustrate evolutionary graphics for streaming data in a printed page. However, we have three basic suggestions: 1) waterfall diagrams, 2) transient geographic mapping, 3) multivariate visualization for port scanning and denial of service attacks.

4.1 Waterfall Diagrams

The waterfall diagram is a staple of sonar signal visualization and can take either of two forms. In either form, for a given (very small) time epoch there is a scan, which marks the source of the signal. The scan may be in azimuth, i.e. 0 to 360 degrees or in frequency, typically 0 to perhaps a few thousand Hertz. The appropriate signal/frequency is marked. At the end of the first epoch, the scan line is dropped down and a new epoch begins repeating the procedure. The procedure is repeated perhaps a thousand times until the screen is filled. Persistent targets in azimuth may be identified with ships. Targets that are moving may be identified by diagonal lines in the waterfall diagram. Similarly the suite of frequencies identified in a spectral waterfall can identify the type of ship and whether rotating machinery is constant in speed or changing. A similar notion can be exploited for Internet traffic data. For example the source IP or destination IP may be plotted in a waterfall diagram as a function of time. We propose to develop these graphics.

4.2 Transient Geographic Mapping

Transient geographic mapping is much harder to illustrate, but comparatively easy to understand. Most users belong to a class B network. Class A are typically reserved for very large providers of backbone services such as AT&T, Sprint, MCI and the like. Thus the first two octets can typically be identified with corporate entities including ISPs, university or government users, whether national or international. International corporations may use the same first two octets in widely geographically distributed regions, but to a large extent the first two octets can be reasonably geographically localized. We suggest two types of transient geographic mapping. The idea is to identify the first two octets with a geographic location, usually the headquarters of the class B network owners. Two forms of transient displays are desirable. First an unthesholded display for which every packet from a source IP lights up the source geographic point with a fairly rapid decay. Thus sessions for which many packets are being sent from a source will have a persistent bright display, with less persistent displays for sources sending fewer packets. This type of display is useful in a benign situation for gathering ground truth average traffic. However, in a denial of service attack, this would be useful for rapidly identifying the sources of the attack.

A second suggestion is to threshold the high frequency traffic and plot only low frequency packets with a long persistence. Characteristically, intruders tend to try to attack systems stealthily so that probing packets are sent infrequently so as not to arouse suspicion. Thus making infrequent packets from a particular pair of octets may be quite useful in identifying would-be intruders.

4.3 Other Plotting Devices

Anomaly detection in Internet traffic can highlight other threats or unusual behavior. Consider for example the drill-down (pan and zoom). Plotting the Number of Bytes versus the Number of Packets can characterize different services. Automated services such as email and ftp will tend to lie along a diagonal, whereas telnet services typically carried out by an individuals keystrokes will tend to have a lot of packets with comparatively few number of bytes, i.e. telnet commands tend to be small. Web services can range over a wide range of combinations of Number of Bytes versus Number of Packets depending on the size of the web page being transferred and the number of images it contains. Another Pan and Zoom that could be effective is plotting the Number of Bytes versus Duration or Source IP versus Destination IP. The vertical banding in this type of illustration indicates many Source IP connections to a single destination IP, not uncommon for a popular destination such as Google, eBay, or Amazon. However, horizontal banding indicates a connection from a single source IP to many destination IPs. This would tend to suggest that the source IP was probing various destination IPs with possibly malicious intent. Of course search engines such as Google are constantly probing web servers for new or updated web pages, so such probings can be comparatively benign. With dynamic evolutionary graphics we could plot destination IP versus source IP versus time in a stereoscopic display.

5. Proposed Task Summary

• I propose what I like to call *evolutionary graphics*, i.e. graphics, which evolve as a function of new data being added. The combination of recursive algorithms and evolutionary graphics will provided a fundamental approach for analyzing streaming data.

• I propose to develop scalable graphics devices (pan and zoom) for discrete data that has many more points than screen resolution would allow.

• I propose to investigate the theoretical properties and implementation of the quantizing algorithm for truly massive streaming data .

• I propose to investigate recursive kernel density estimators in the critical multidimensional case.

• I propose to investigate adaptive mixtures density estimation algorithms using orthonormal bases such as wavelets to reduce or eliminate spurious terms.

• I propose to investigate classical exponential smoother with adaptive time scaling and their potential use as multiscale data analysis.

• I propose to develop evolutionary graphics tools including waterfall diagrams, transient geographic mapping, and related multivariate pan and zoom methodology.

References

Committee on Network-Centric Naval Forces, Naval Studies Board, National Research Council (2000) *Network-Centric Naval Forces: A Transition Strategy for Enhancing Operational Capabilities* (512 pp.), Washington, DC: National Academy of Science

Khumbah, N.-A. and Wegman, E. J. (2003) "Data compression by geometric quantization," to appear in *Recent Advances and Trends in Nonparametric Statistics*, (M. Akritas and D.N. Politis, eds.), Amsterdam: North Holland Elsevier

Marchette, D. J. (2001) *Computer Intrusion Detection and Network Monitoring: A Statistical Viewpoint*, New York: Springer-Verlag

Marchette, D. J. and Wegman, E. J. (2003) "Statistical analysis of network data for cybersecurity," to appear *Chance*

Priebe, C. E. (1994) "Adaptive mixtures," *Journal of the American Statistical Association*, 89, 796-806

Wegman, E. J. and Davies, H. I. (1979) "Remarks on some recursive estimators of a probability density," *Annals of Statistics*, 7, 316-327

Wolverton, C. T. and Wagner, T. J. (1969) "Asymptotically optimal discriminant functions for pattern classification," *IEEE Transactions on Information Theory*, IT-15, 258-265.

Yamato, H. (1971) "Sequential estimation of a continuous probability density function and the mode," *Bulletin of Mathematical Statistics*, 14, 1-12.

Biosketch of Edward J. Wegman

Professor Edward J. Wegman received his B.S. in mathematics degree from St. Louis University in 1965. He received the M.S. and Ph.D. degrees in mathematical statistics from the University of Iowa, the latter degree in 1968. Subsequently, he spent 10 years on the faculty of the world-class Department of Statistics at the University of North Carolina. Dr. Wegman's early career focused on the development of aspects of the theory of mathematical statistics. In 1978, Professor Wegman went to the Office of Naval Research (ONR) where he was the Head of the Mathematical Sciences Division. In this role, he had responsibility Navy-wide for basic research programs in applied mathematics, statistics and probability, systems theory, operations research, discrete mathematics, communication theory, and numerical analysis and computational architectures. In addition, he was responsible for a variety of cross-disciplinary areas including such projects as mathematical models of biological intelligence, mathematical methods for remote sensing, and topological methods in chemistry. As part of his duties at the Office of Naval Research, coined the phrase, computational statistics, and developed a high profile research area around this concept. The idea was to focus on techniques and methodologies that could not be achieved without the capabilities of modern computing resources. This program led to a revolution in contemporary statistical graphics. Dr. Wegman was the original program director of the basic research program in Ultra High Speed Computing at the Strategic Defense Initiative's Innovative Science and Technology Office (Star Wars Program). As the SDI program officer, Dr. Wegman was responsible for programs in software development tools, highly parallel architectures and optical computing.

Dr. Wegman came to George Mason University with an extensive background in both theoretical statistics and computing technology, with an extensive knowledge of the considerable data analytic problems associated with large scale scientific and technical databases and with a strong motivation to develop the computational and methodological tools to address these problems. In 1986, he launched the Center for Computational Statistics and developed the M.S. in Statistical Science degree program. More recently he has been involved with the development of the School of Computational Sciences and the Ph.D. program in Computational Sciences and Informatics at George Mason University. He holds a joint appointment in the School of Information Technology and Engineering and the School of Computational Sciences.

He has been consultant to a variety of governmental and private sector organizations including the National Research Council of the National Academy of Science, the states of North Carolina and Ohio, the U.S. Navy and the Executive Office of Management and Budget. He has just finished serving a two-year term on the NRC Committee for Naval Force's Capability for Theater Missile Defense. He has organized some twenty major workshops and conferences, including both the 1988 and the 2002 Symposia on the Interface of Computing Science and Statistics. He has served as associate editor of the *Journal of the American Statistical Association*, *Statistics and Probability Letter*s and *Communications in Statistics*. He presently serves on the editorial boards of the *Journal of Statistical Planning and Inference*, the *Naval Research Logistics Quarterly*, the *Journal of Nonparametric Statistics* and *Computational Statistics and Data Analysis*. Dr. Wegman completed a four-year term as the Theory and Methods editor of the prestigious *Journal of the American Statistical Association*. He is the founder of the Interface Foundation of North America, Inc., which is the host organization for the Symposia on the Interface of Computing Science and Statistics. The Interface Foundation in conjunction with the

American Statistical Association and the Institute for Mathematical Statistics has also launched the interdisciplinary *Journal of Computational and Graphical Statistics*. Dr. Wegman served in national office in the Institute of Mathematical Statistics, the American Statistical Association and the American Association for the Advancement of Science. He has published more than 160 papers and seven books. His professional stature has been recognized by his election as Fellow of the American Statistical Association, the American Association for the Advancement of Science, the Washington Academy of Science, and the Institute of Mathematical Statistics. In addition he was elected as a Senior Member of IEEE. Dr. Wegman has been elected to membership in the International Statistical Institute. Dr. Wegman has also received numerous Navy awards including the Navy's Meritorious Civilian Service Medal. Wegman has received the 1990 Distinguished Faculty Award, the 1999 Outstanding Research Award from George Mason University, the 1999 Army Wilks Medal from the U.S. Army and the 2002 ASA Founder's Award. Dr. Wegman came to George Mason University in 1986 and is the Bernard J. Dunn Professor of Information Technology and Applied Statistics and the Director of the Center for Computational Statistics. He was the Founding Chair of the Department of Applied and Engineering Statistics. A full CV may be found at

http://www.galaxy.gmu.edu/stats/faculty/wegman.resume2.html.

CURRENT AND PENDING SUPPORT

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

Investigator: Dr. Edward Wegman

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

Project/Proposal Title: Intrusion Detection Using Data Mining Techniques

Source of Support: Air Force Office of Scientific Research (AFOSR)
Award Amount (or Annual Rate:) $ 470,275 Period Covered: 4/15/2001-10/14/2003
Location of Research: George Mason University
Person-Months Committed to the Project: Cal: Acad: 0.00 Summer: 0.00

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

Project/Proposal Title: Adaptive Data Cube for Integrated Sensing and Processing

Source of Support: Johns Hopkins University (Prime: DARPA)
Award Amount (or Annual Rate:) $ 450,000 Period Covered: 7/1/2001-6/30/2004
Location of Research: George Mason University
Person-Months Committed to the Project: Cal: Acad: 1.30 Summer: 3.00

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

Project/Proposal Title: Analytic and Graphical Methods for Streaming Data with Applications to Netcentric Warfare

Source of Support: Office of Naval Research
Award Amount (or Annual Rate:) $ 284,601 Period Covered: 10/16/2003-10/15/2006
Location of Research: George Mason University
Person-Months Committed to the Project: Cal: Acad: Summer: 2.00

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

Project/Proposal Title:

Source of Support:
Award Amount (or Annual Rate:) $ Period Covered:
Location of Research:
Person-Months Committed to the Project: Cal: Acad: Summer:

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

Project/Proposal Title:

Source of Support:
Award Amount (or Annual Rate:) $ Period Covered:
Location of Research:
Person-Months Committed to the Project: Cal: Acad: Summer:

* 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, EQUIPMENT, AND OTHER RESOURCES

FACILITIES: Identify the facilities to be used at each performance site listed and, as appropriate, briefly indicate their capabilities, pertinent capabilities, relative proximity, and extent of availability to the project. Use "Other" to describe the facilities at any other performance sites listed and at sites for field studies. Use additional pages, if necessary.

The Center for Computational Statistics has a number of relevant facilities. In addition to a Virtual Reality Lab with Silicon Graphics Onyx II machines driving CRT projection systems, The Center has recently created a PC-based MiniCAVE system with 16:9 aspect ratio high-definition LCD projectors operating with polarized light stereo mode. The Center, in conjunction with GMU's CIO is collecting data on all IP header traffic in and out of the university and has capability of storing 5.4 terabytes of data.

MAJOR EQUIPMENT: List the most important items available for this project and, as appropriate, identify the location and pertinent capabilities of each:

ADDITIONAL INFORMATION: Provide any other information describing the other resources available for the project. Identify support services such as consultant, secretarial, machine shop, and electronics shop, and the extent to which they will be available for the project. Include an explanation of any consortium/contractual/subaward arrangements with other organizations.

45810

George Mason University

Budget
(Proposal Section)

Fairfax, Virginia 22030-4444

September 23, 2004

Dr. Robert L. Launer
US Army Research Office

Dear Dr. Launer:

Enclosed please find a budget revision for the proposal submitted to the Army Research Office for Dr. Edward Wegman, Center for Computational Statistics (CCS), George Mason University. The proposal is entitled **Analytical and Graphical Methods for Streaming Data with Applications to Netcentric Warfare.**

If you have any questions regarding the technical content of this project, please feel free to contact Dr. Wegman at 703/993-1691 or ewegman@gmu.edu. Questions regarding budget, university policies and procedures should be directed to Patricia Carcamo, Grants Administrator, Office of Sponsored Programs at 703/993-2987.

Sincerely,

Ann T. McGuigan, Ph.D.
Director
Office of Sponsored Programs

Enclosures

cc: E. Wegman

George Mason University
4400 University Drive
Fairfax, VA 22030

Army Research Office (ARO)
10/16/2003-10/15/2006

		YEAR ONE	GMU IN-KIND	YEAR TWO	GMU IN-KIND	YEAR THREE	GMU IN-KIND	TOTAL
A. PERSONNEL								
1. Faculty - Academic Year	FTE							
PI: Dr. Ed Wegman	0.00	0	0	0	0	0	0	0
2. Faculty - Summer/Part time	FTE							
PI: Dr. Ed Wegman	0.22	34,010	0	35,711	0	37,497	0	107,218
3. Students (AY)	No.							
GRA - Doctoral $16.67/Hr X 359.9 Hrs	1	6,000	0	6,000	0	6,000	0	18,000
Students (SUM)								
GRA - Doctoral	0	0	0	0	0	0	0	0
TOTAL PERSONNEL		**40,010**	**0**	**41,711**	**0**	**43,497**	**0**	**125,218**
B. FRINGE BENEFITS								
@ 24.14%		0	0	0	0	0	0	0
@ 7.65%		3,061	0	3,191	0	3,328	0	9,580
TOTAL FRINGE		**3,061**	**0**	**3,191**	**0**	**3,328**	**0**	**9,580**
C. TRAVEL								
1. Domestic Travel 1 Trip/Yr		2,366	0	1,801	0	1,829	0	5,996
TOTAL TRAVEL		**2,366**	**0**	**1,801**	**0**	**1,829**	**0**	**5,996**
D. EQUIPMENT								
1. Computers		0	0	0	0	0	0	0
TOTAL EQUIPMENT		**0**	**0**	**0**	**0**	**0**	**0**	**0**
E. OTHER DIRECT COSTS								
1. Tuition Doctoral GRA's								
- Out/State: 6 Credit Hrs @ $686/Hr		4,116	0	4,281	0	4,452	0	12,849
TOTAL OTHER		**4,116**	**0**	**4,281**	**0**	**4,452**	**0**	**12,849**
TOTAL DIRECT COSTS		**49,553**	**0**	**50,984**	**0**	**53,106**	**0**	**153,643**
F. INDIRECT COSTS								
45% of Modified Total Direct Costs		20,447	0	21,016	0	21,894	0	63,358
Provisional Inidrect Rate as of 07/01/2004								
TOTAL COSTS		**70,000**	**0**	**72,000**	**0**	**75,000**	**0**	**217,000**

Percent of GMU Cost Sharing 0.00%

TOTAL DIRECT COSTS REQUESTED FROM SPONSOR	153,643
TOTAL INDIRECT COSTS REQUESTED FROM SPONSOR	63,358
TOTAL COSTS REQUESTED FROM SPONSOR	**217,000**
TOTAL GMU IN-KIND CONTRIBUTION	0
TOTAL PROGRAM COSTS	**217,000**

Salaries and wages are estimates only. Actual salaries and wages will be paid in accordance with University policy.
Tuition and fees are budgeted at Out/State rates. Actual charges will be made according to individual domicilliary classification.

Budget preparation: S:BlueTeam/IT&E/AES/WegmanARO7-03

Patricia Carcamo

24-Jul-03
Patricia M. Carcamo

SUMMARY

PROPOSAL BUDGET

YEAR- 1

OFFEROR George Mason University Ann T. McGuigan, Ph.D.						
PRINCIPAL INVESTIGATOR/PROJECT DIRECTOR (PI/PD) Dr. Edward Wegman						
A. SENIOR PERSONNEL, PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title. A 7. show number in parentheses)	Man Hrs/Mo	Rates	-Mos CAL	ACAD	SMR	Funds Requested by Offeror
1. PI: Dr. Edward Wegman	181.81	93.53			2.00	34,010
2.						
3.						
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET EXPLANATION)						
7. ( ) TOTAL SENIOR PERSONNEL (1-6)						34,010
B. OTHER PERSONNEL (SHOW NUMBERS IN PARENTHESES)						
1. ( ) POST DOCTORAL ASSOCIATES						
2. ( ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)						
3. (1) GRADUATE STUDENTS						13,000
4. ( ) UNDERGRADUATE STUDENTS						
5. ( ) SECRETARIAL - CLERICAL (If charged directly)						
6. ( ) OTHER						
7. TOTAL SALARIES AND WAGES (A + B)						47,010
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						3596
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						50,606
D. PERMANENT EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.00. ATTACH ADDITIONAL EXPLANATION PAGES, IF NECESSARY.)						
TOTAL PERMANENT EQUIPMENT						
E. TRAVEL (LIST ON BUDGET EXPLANATION PAGE)						
1. DOMESTIC (INCLUDE CANADA, MEXICO, AND U.S. POSSESSIONS)						2,500
2. FOREIGN						
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS $						
2. TRAVEL						
3. SUBSISTENCE						
4. OTHER						
( ) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS (ITEMIZE ON BUDGET EXPLANATION PAGE)						
1. MATERIALS AND SUPPLIES						
2. PUBLICATIONS COSTS/DOCUMENTATION/DISSEMINATION						
3. CONSULTANT SERVICES						
4. COMPUTER (ADPE) SERVICES						
5. SUBAWARDS						
6. OTHER						8,232
7. TOTAL OTHER DIRECT COSTS						8,232
H. TOTAL DIRECT COSTS (A THROUGH G)						61,338
I. INDIRECT COSTS		Rate	Base	Total		
	Overhead	48.00%	53,106	25,491		
	G&A	0.00%		0		
	Fringe	0.00%		0		
TOTAL INDIRECT COSTS	FCCM	0.00000%		0		25,491
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						86,829
K. FEE (%) BASE $						
L. COST SHARING						
M. AMOUNT OF THIS REQUEST						86,829
PI/PD NAME (TYPED) & SIGNATURE Dr. Edward Wegman Edward J. Wegman						DATE 7/29/03
OFFEROR'S AUTHORIZED REP. NAME (TYPED) & SIGNATURE George Mason University Ann T. McGuigan, Ph.D. Ann T. McGuigan						DATE 7-29-03

SUMMARY

PROPOSAL BUDGET

YEAR- 2

OFFEROR	George Mason University Dr. Ann T. McGuigan, Ph.D.					
PRINCIPAL INVESTIGATOR/PROJECT DIRECTOR (PI/PD)	Dr. Edward Wegman					
A. SENIOR PERSONNEL, PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title. A 7. show number in parentheses)	Man Hrs/Mo	Rates	-Mos CAL	ACAD	SMR	Funds Requested by Offeror
1. PI: Dr. Edward Wegman	181.81	98.21			2.00	35,711
2.						
3.						
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET EXPLANATION)						
7. ( ) TOTAL SENIOR PERSONNEL (1-6)						35,711
B. OTHER PERSONNEL (SHOW NUMBERS IN PARENTHESES)						
1. ( ) POST DOCTORAL ASSOCIATES						
2. ( ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)						
3. (1) GRADUATE STUDENTS						13,650
4. ( ) UNDERGRADUATE STUDENTS						
5. ( ) SECRETARIAL - CLERICAL (If charged directly)						
6. ( ) OTHER						
7. TOTAL SALARIES AND WAGES (A + B)						49,361
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						3776
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						53,137
D. PERMANENT EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.00. ATTACH ADDITIONAL EXPLANATION PAGES, IF NECESSARY.)						
TOTAL PERMANENT EQUIPMENT						
E. TRAVEL (LIST ON BUDGET EXPLANATION PAGE)						
1. DOMESTIC (INCLUDE CANADA, MEXICO, AND U.S. POSSESSIONS)						2,500
2. FOREIGN						
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS $						
2. TRAVEL						
3. SUBSISTENCE						
4. OTHER						
( ) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS (ITEMIZE ON BUDGET EXPLANATION PAGE)						
1. MATERIALS AND SUPPLIES						
2. PUBLICATIONS COSTS/DOCUMENTATION/DISSEMINATION						
3. CONSULTANT SERVICES						
4. COMPUTER (ADPE) SERVICES						
5. SUBAWARDS						8,891
6. OTHER						8,891
7. TOTAL OTHER DIRECT COSTS						64,528
H. TOTAL DIRECT COSTS (A THROUGH G)						
I. INDIRECT COSTS		Rate	Base	Total		
	Overhead	48.00%	55,637	26,706		
	G&A	0.00%		0		
	Fringe	0.00%		0		
	FCCM	0.00000%		0		
TOTAL INDIRECT COSTS						26,706
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						91,234
K. FEE (%) BASE $						
L. COST SHARING						
M. AMOUNT OF THIS REQUEST						91,234
PI/PD NAME (TYPED) & SIGNATURE Dr. Edward Wegman	Edward J. Wegman				DATE	7/29/03
OFFEROR'S AUTHORIZED REP. NAME (TYPED) & SIGNATURE George Mason University Dr. Ann T. McGuigan, Ph.D	Ann T. McGuigan				DATE	7-29-03

SUMMARY

PROPOSAL BUDGET

YEAR- 3

OFFEROR	George Mason University Dr. Ann T. McGuigan, Ph.D.					
PRINCIPAL INVESTIGATOR/PROJECT DIRECTOR (PI/PD)	Dr. Edward Wegman					
A. SENIOR PERSONNEL, PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title. A 7 show number in parentheses)	Man Hrs/Mo	Rates	-Mos CAL	-Mos ACAD	-Mos SMR	Funds Requested by Offeror
1. PI: Dr. Edward Wegman	181.81	103.12			2.00	37,497
2.						
3.						
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET EXPLANATION)						
7. ( ) TOTAL SENIOR PERSONNEL (1-6)						37,497
B. OTHER PERSONNEL (SHOW NUMBERS IN PARENTHESES)						
1. ( ) POST DOCTORAL ASSOCIATES						
2. ( ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)						
3. (1) GRADUATE STUDENTS						14,333
4. ( ) UNDERGRADUATE STUDENTS						
5. ( ) SECRETARIAL - CLERICAL (If charged directly)						
6. ( ) OTHER						
7. TOTAL SALARIES AND WAGES (A + B)						51,830
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						3965
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						55,795
D. PERMANENT EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.00. ATTACH ADDITIONAL EXPLANATION PAGES, IF NECESSARY.)						
TOTAL PERMANENT EQUIPMENT						0
E. TRAVEL (LIST ON BUDGET EXPLANATION PAGE)						
1. DOMESTIC (INCLUDE CANADA, MEXICO, AND U.S. POSSESSIONS)						2,500
2. FOREIGN						
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS $						
2. TRAVEL						
3. SUBSISTENCE						
4. OTHER						
( ) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS (ITEMIZE ON BUDGET EXPLANATION PAGE)						
1. MATERIALS AND SUPPLIES						
2. PUBLICATIONS COSTS/DOCUMENTATION/DISSEMINATION						
3. CONSULTANT SERVICES						
4. COMPUTER (ADPE) SERVICES						
5. SUBAWARDS						
6. OTHER						9,602
7. TOTAL OTHER DIRECT COSTS						9,602
H. TOTAL DIRECT COSTS (A THROUGH G)						67,897
I. INDIRECT COSTS		Rate	Base	Total		
	Overhead	48.00%	58,295	27,982		
	G&A	0.00%		0		
	Fringe	0.00%		0		
TOTAL INDIRECT COSTS	FCCM	0.00000%		0		27,982
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						95,879
K. FEE (%) BASE $						
L. COST SHARING						
M. AMOUNT OF THIS REQUEST						95,879
PI/PD NAME (TYPED) & SIGNATURE Dr. Edward Wegman	Edward J. Wegman				DATE	7/29/03
OFFEROR'S AUTHORIZED REP. NAME (TYPED) & SIGNATURE George Mason University Dr. Ann T. McGuigan, Ph.D	Ann T. McGuigan				DATE	7-29-03

SUMMARY

PROPOSAL BUDGET

YEAR-

OFFEROR George Mason University Ann T. McGuigan, Ph.D.

PRINCIPAL INVESTIGATOR/PROJECT DIRECTOR (PI/PD) Dr. Edward Wegman

A. SENIOR PERSONNEL, PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title. A 7. show number in parentheses)	Man Hrs/Mo	Rates	-Mos CAL	-Mos ACAD	-Mos SMR	Funds Requested by Offeror
1. PI: Dr. Edward Wegman	181.81	98.29			6.00	107,218
2.						
3.						
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET EXPLANATION)						
7. ( ) TOTAL SENIOR PERSONNEL (1-6)						107,218
B. OTHER PERSONNEL (SHOW NUMBERS IN PARENTHESES)						
1. ( ) POST DOCTORAL ASSOCIATES						
2. ( ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)						
3. (1) GRADUATE STUDENTS						40,983
4. ( ) UNDERGRADUATE STUDENTS						
5. ( ) SECRETARIAL - CLERICAL (If charged directly)						
6. ( ) OTHER						
7. TOTAL SALARIES AND WAGES (A + B)						148,201
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						11337
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						159,538
D. PERMANENT EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.00. ATTACH ADDITIONAL EXPLANATION PAGES, IF NECESSARY.)						
TOTAL PERMANENT EQUIPMENT						
E. TRAVEL (LIST ON BUDGET EXPLANATION PAGE)						7,500
1. DOMESTIC (INCLUDE CANADA, MEXICO, AND U.S. POSSESSIONS)						
2. FOREIGN						
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS $						
2. TRAVEL						
3. SUBSISTENCE						
4. OTHER						
( ) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS (ITEMIZE ON BUDGET EXPLANATION PAGE)						
1. MATERIALS AND SUPPLIES						
2. PUBLICATIONS COSTS/DOCUMENTATION/DISSEMINATION						
3. CONSULTANT SERVICES						
4. COMPUTER (ADPE) SERVICES						
5. SUBAWARDS						26,725
6. OTHER						26,725
7. TOTAL OTHER DIRECT COSTS						193,763
H. TOTAL DIRECT COSTS (A THROUGH G)						

I. INDIRECT COSTS		Rate	Base	Total	
	Overhead	48.00%	167,038	80,178	
	G&A	0.00%		0	
	Fringe	0.00%		0	
TOTAL INDIRECT COSTS	FCCM	0.00000%		0	80,178
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					273,941
K. FEE (%) BASE $					
L. COST SHARING					
M. AMOUNT OF THIS REQUEST					273,941

PI/PD NAME (TYPED) & SIGNATURE
Dr. Edward Wegman Edward J. Wegman DATE 7/29/03

OFFEROR'S AUTHORIZED REP. NAME (TYPED) & SIGNATURE
George Mason University Ann T. McGuigan, Ph.D. Ann T. McGuigan DATE 7-29-03

George Mason University
4400 University Drive
Fairfax, VA 22030

Army Research Office (ARO)
10/16/2003-10/15/2006

		YEAR ONE	GMU IN-KIND	YEAR TWO	GMU IN-KIND	YEAR THREE	GMU IN-KIND	TOTAL
A. PERSONNEL								
1. Faculty - Academic Year	FTE							
PI: Dr. Ed Wegman	0.00	0	0	0	0	0	0	0
2. Faculty - Summer/Part time	FTE							
PI. Dr. Ed Wegman	0.22	34,010	0	35,711	0	37,497	0	107,218
3. Students (AY)	No.							
GRA - Doctoral $16.67 p/h X 780 hrs	1	13,000	0	13,650	0	14,333	0	40,983
Students (SUM)								
GRA - Doctoral	0	0	0	0	0	0	0	0
TOTAL PERSONNEL		**47,010**	**0**	**49,361**	**0**	**51,830**	**0**	**148,201**
B. FRINGE BENEFITS								
@ 23.46%		0	0	0	0	0	0	0
@ 7.65%		3,596	0	3,776	0	3,965	0	11,337
TOTAL FRINGE		**3,596**	**0**	**3,776**	**0**	**3,965**	**0**	**11,337**
C. TRAVEL								
1. Domestic travel 2 trips/yr		2,500	0	2,500	0	2,500	0	7,500
TOTAL TRAVEL		**2,500**	**0**	**2,500**	**0**	**2,500**	**0**	**7,500**
D. EQUIPMENT								
1. Computers		0	0	0	0	0	0	0
TOTAL EQUIPMENT		**0**	**0**	**0**	**0**	**0**	**0**	**0**
E. OTHER DIRECT COSTS								
1. Tuition Doctoral GRA's								
- Out/State: 12 Credit Hrs @ $686/Hr		8,232	0	8,891	0	9,602	0	26,725
TOTAL OTHER		**8,232**	**0**	**8,891**	**0**	**9,602**	**0**	**26,725**
TOTAL DIRECT COSTS		**61,338**	**0**	**64,528**	**0**	**67,897**	**0**	**193,763**
F. INDIRECT COSTS								
48% of Modified Total Direct Costs		25,491	0	26,706	0	27,982	0	80,178
TOTAL COSTS		**86,829**	**0**	**91,234**	**0**	**95,879**	**0**	**273,941**
Percent of GMU Cost Sharing		0.00%						

TOTAL DIRECT COSTS REQUESTED FROM SPONSOR	193,763
TOTAL INDIRECT COSTS REQUESTED FROM SPONSOR	80,178
TOTAL COSTS REQUESTED FROM SPONSOR	273,941
TOTAL GMU IN-KIND CONTRIBUTION	0
TOTAL PROGRAM COSTS	273,941

Salaries and wages are estimates only. Actual salaries and wages will be paid in accordance with University policy.
Tuition and fees are budgeted at Out/State rates. Actual charges will be made according to individual domicilliary classification.
GMU's negotiated facilities and administrative rates (F&A) expire June 30, 2004. Project costs beyond that date are estimates subject to revision when new F&A rates are approved.

Budget preparation: S:BlueTeam/IT&E/AES/WegmanARO7-03

Lauren P. Magruder
24-Jul-03
Lauren P. Magruder

CERTIFICATIONS
EDUCATIONAL/NONPROFIT INSTITUTIONS

The following certifications apply to the proposal titled: Analytic and Graphical Methods for Streaming Data with Applications to Netcentric Warfare

A. 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.

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 an 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 Form to Reporting Lobbying," 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.

B. CERTIFICATION REGARDING DEBARMENT, SUSPENSION, AND OTHER RESPONSIBILITY MATTERS--PRIMARY COVERED TRANSACTIONS

(1) The prospective primary participant certifies, to the best of its knowledge and belief, that it and its principals:

(a) Are not presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded by any Federal department or agency;

(b) Have not within a three-year period preceding this proposal been convicted of or had a civil judgment rendered against them for commission of fraud or a criminal offense in connection with obtaining, attempting to obtain, or performing a public (Federal, State or local) transaction or contract under a public transaction; violation of Federal or State antitrust statutes or commission of embezzlement, theft, forgery, bribery, falsification or destruction of records, making false statements, or receiving stolen property;

(c) Are not presently indicted for or otherwise criminally or civilly charged by a government entity (Federal, State or local) with commission of any of the offenses enumerated in paragraph (1)(b) of this certification; and

(d) Have not within a three-year period preceding this application/proposal had one or more public transactions (Federal, State or local) terminated for cause or default.

(2) Where the prospective primary participant is unable to certify to any of the statements in this certification, such prospective participant shall attach an explanation to this proposal.

C. CERTIFICATION REGARDING DRUG-FREE WORKPLACE REQUIREMENTS

A. The grantee certifies that it will or will continue to provide a drug-free workplace by:

(a) Publishing a statement notifying employees that the unlawful manufacture, distribution, dispensing, possession or use of a controlled substance is prohibited in the grantee's workplace and specifying the actions that will be taken against employees for violation of such prohibition;

(b) Establishing an on-going drug-free awareness program to inform employees about--

(1) The dangers of drug abuse in the workplace;
(2) The grantee's policy of maintaining a drug-free workplace;
(3) Any available drug counseling, rehabilitation, and employee assistance programs; and
(4) The penalties that may be imposed upon employees for drug abuse violations occurring in the workplace;

(c) Making it a requirement that each employee to be engaged in the performance of the grant be given a copy of the statement required by paragraph (a);

(d) Notifying the employee in the statement required by paragraph (a) that, as a condition of employment under the grant, the employee will--

(1) Abide by the terms of the statement; and

(2) Notify the employer in writing of his or her conviction for a violation of a criminal drug statute occurring in the workplace no later than five calendar days after each conviction;

(e) Notifying the agency in writing, within ten calendar days after receiving notice under subparagraph (d)(2) from an employee or otherwise receiving actual notice of such conviction. Employers of convicted employees must provide notice, including position title, to every grant officer or other designee on whose grant activity the convicted employee was working, unless the Federal agency has designated a central point for the receipt of such notices. Notice shall include the identification number(s) of each affected grant;

(f) Taking one of the following actions, within 30 calendar days of receiving notice under subparagraph (d)(2), with respect to any employee who is so convicted--

(1) Taking appropriate personnel action against such an employee, up to and including termination, consistent with the requirements of the Rehabilitation Act of 1973, as amended; or

(2) Requiring such employee to participate satisfactorily in a drug abuse assistance or rehabilitation program approved for such purposes by a Federal, State, or local health, law enforcement, or other appropriate agency;

(g) Making a good faith effort to continue to maintain a drug-free workplace through implementation of paragraphs (a), (b), (c), (d), (e), and (f).

B. The grantee may insert in the space provided below the site(s) for the performance of work done in connection with the specific grant:

Place of Performance (street address, city, county, state, zip code)
4400 University Drive
Fairfax, Fairfax County, VA 22030

Check (__) if there are workplaces on file that are not identified here.

ORGANIZATION: George Mason University
AUTHORIZED REPRESENTATIVE SIGNATURE: Ann J. McGuigan DATE: 7-29-03
TYPED NAME AND TITLE: Ann T. McGuigan, Ph.D. Director, Office of Sponsored Pgms.
TELEPHONE NUMBER: 703-993-2988 FAX NUMBER: 703-993-2296
E-MAIL: amcguiga@gmu.edu

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
800 NORTH QUINCY STREET
ARLINGTON, VA 22217-5660

IN REPLY REFER TO

NEGOTIATION AGREEMENT

INSTITUTION: GEORGE MASON UNIVERSITY
4400 University Drive
Fairfax, VA 22030-4444

The Facilities and Administrative (F&A) cost rates contained herein are for use on all grants and/or contracts issued or awarded to George Mason University by all Federal agencies of the United States of America in accordance with the cost principles mandated by the Office of Management and Budget (OMB) Circular A-21. These rates shall be used for forward pricing and billing purposes for George Mason University's Fiscal Years (FYs) 2002-2004. This rate agreement supersedes all previous rate agreements/determinations for FYs 2002-2004.

SECTION I: RATES - TYPE: PREDETERMINED (PRED)

	From	To	Rate	Location	Base	Applicable To
Pred.	7/1/01	6/30/03	45.5%	On Campus	(a)	Organized Research (1)
Pred.	7/1/03	6/30/04	45.1%	On Campus	(a)	Organized Research (1)
	7/1/01	6/30/04	26.0%	Off Campus	(a)	Organized Research (1)
Pred.	7/1/01	6/30/03	48.4%	On Campus	(a)	Organized Research (2)
Pred.	7/1/03	6/30/04	48.0%	On Campus	(a)	Organized Research (2)
	7/1/01	6/30/04	28.9%	Off Campus	(a)	Organized Research (2)

DISTRIBUTION BASE

(a) Modified total direct costs, consisting of all salaries and wages, fringe benefits, materials and supplies, services, travel, and subgrants and subcontracts up to the first $25,000 of each subgrant or subcontract (regardless of the period covered by the subgrant or subcontract). Equipment, capital expenditures, charges for patient care and tuition remission, rental costs, scholarships, and fellowships as well as the portion of each subgrant and subcontract in excess of $25,000 shall be excluded from modified total direct costs.

APPLICABLE:

(1) Applies to all Non-DOD Instruments, all DOD Grants and to DOD Contracts awarded before November 30, 1993.

(2) Applies to only DOD Contracts awarded on or after November 30, 1993 in accordance with and under the authority of DFARS 231.303(1). See Section II, Paragraph E hereof.

SECTION II - GENERAL TERMS AND CONDITIONS

A. **LIMITATIONS:** Use of the rates set forth under Section I of this agreement is expressly subject to any statutory or administrative limitations and is applicable to a given grant, contract, or other agreement only to the extent that funds are available and consistent with any and all limitations of cost clauses or provisions, if any, contained therein. Acceptance of any or all of the rates agreed to herein is predicated upon all the following conditions: (1) that no costs other than those incurred by the grantee/contractor were included in its indirect cost pool as finally accepted and that all such costs are legal obligations of the grantee/contractor and allowable under the governing and applicable cost principles; (2) that the same costs that have been treated as indirect costs are not claimed as direct costs; (3) that similar types of costs have been accorded consistent accounting treatment; and (4) that the information provided by the contractor/grantee, which was used as the basis for the acceptance of the rate(s) agreed to herein and expressly relied upon by the Government in negotiating and accepting the said rates, is not subsequently found to be materially incomplete or inaccurate.

B. **ACCOUNTING CHANGES:** The predetermined rates contained in Section I of this agreement are based on the accounting system in effect at the time this agreement was negotiated. Changes to the method(s) of accounting for costs which affects the amount of reimbursement resulting from the use of these rates requires the express and written approval of the authorized representative of the cognizant negotiating agency for the Government prior to implementation of any such changes. Such changes include but are not limited to changes in the charging of a particular type of cost from indirect to direct. Failure on the part of the grantee/contractor to obtain the required approval may result in subsequent cost disallowances.

C. **PREDETERMINED RATES:** The predetermined rates contained in this agreement are not subject to adjustment in accordance with the provisions of OMB Circular A-21, subject to the limitations contained in Part A of this section.

D. **USE BY OTHER FEDERAL AGENCIES:** The rates set forth in Section I hereof were negotiated in accordance with and under the authority set forth in OMB Circular A-21. Accordingly, such rates shall be applied, to the extent provided in such circular, to grants and contracts to which OMB Circular A-21 is applicable, subject to any limitations in part A of this section. Copies of this document may be provided by either party to other Federal agencies which have or intend to issue or award grants and/or contracts to George Mason University to use the stated rates or to otherwise provide such agencies with documentary notice of this agreement and its terms and conditions.

B. **SPECIAL REMARKS**

APPLICATION OF INDIRECT COST RATES TO DOD CONTRACTS/SUBCONTRACTS: In accordance with DFARS 231.303, no limitation (unless waived by the institution) may be placed on the reimbursement of otherwise allowable indirect costs incurred by an institution of higher education under a DOD contract awarded on or after November 30, 1993, unless the same limitation is applied uniformly to all other organizations performing similar work. It has been determined by the Department of Defense that such limitation is not being uniformly applied. Accordingly, the rates cited (2) of Section I, as explained under the title "APPLICABLE TO" do not reflect the application of the 26% limitation on administrative indirect costs imposed by OMB Circular A-21 whereas (1) does so.

F. **QUALIFICATION**

The purpose of this agreement is to establish F & A cost rates for the period 1 July 2001 through 30 June 2004. These rates are based on GMU's proposal dated January 31, 2001. The government's agreement to the rates in Section I, based on the accounting systems and allocation methodologies in effect at the time this agreement was negotiated, does not represent an agreement that these systems and methodologies are accepted for FY's 2002, 2003, 2004, or any subsequent fiscal year's rate negotiations.

FOR GEORGE MASON UNIVERSITY:

By: Maurice W Scherrens

Maurice W. Scherrens
Senior Vice President

Date: 8/31/01

FOR THE GOVERNMENT:

By: Deborah K. Rafi

Deborah K. Rafi
Contracting Officer

Date: 9-4-01

For information concerning this agreement contact:

Linda B. Shipp (shippl@onr.navy.mil)
Office of Naval Research, Indirect Cost Branch/ONR 242
800 N. Quincy Street, BCT # 1, Room 704
Arlington, VA 22217
(703) 696-8559, FAX: (703) 696-4430