9110-9F

DEPARTMENT OF HOMELAND SECURITY

Science and Technology (S&T) Directorate

[Docket No. DHS-2010-0043]

Agency Information Collection Activities: Submission for Review; Information Collection Request for the Department of Homeland Security (DHS) Science and Technology Protected Repository for the Defense of Infrastructure against Cyber Threats (PREDICT) program.

AGENCY: Science and Technology Directorate, DHS.

ACTION: 60-day Notice and request for comment**.**

SUMMARY: The Department of Homeland Security invites the general public to comment on data collection forms for the Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT) initiative. PREDICT is an initiative to facilitate the accessibility of computer and network operational data for use in cybersecurity defensive research and development. Specifically, PREDICT provides developers and evaluators with regularly updated network operations data sources relevant to cybersecurity defense technology development. The data sets are intended to provide developers with timely and detailed insight into cyberattack phenomena occurring across the Internet and in some cases will reveal the effects of these attacks on networks that are owned or managed by the data producers. A key motivation of PREDICT is to make these data sources more widely available to technology developers and evaluators, who today often determine the efficacy of their technical solutions on anecdotal evidence or small-scale test experiments, rather than on more comprehensive real-world data. The PREDICT website http://www.predict.org/ contains an overview and general information as background, along with the data repository. As specified on the website, access to the PREDICT data repository is available to eligible research groups upon approval of their applications. In addition to helping to determine

whether a group is eligible to access the repository, the forms will also manage the interactions between the PREDICT portal administrators and the research groups accessing the PREDICT portal. The Department is committed to improving its PREDICT initiative and invites interested persons to comment on the following forms and instructions (hereinafter "Forms Package") for the PREDICT initiative: 1) Account Request Form (DHS Form 10029 (12/07)); 2) Request a Dataset Form (DHS Form 10032 (12/07)); 3) My Datasets Form (DHS Form 10033 (12/07)); 4) Memorandum of Agreement - PREDICT (PCC) Coordinating Center and Researcher/User (DHS Form 10035 (12/07)); 5) Memorandum of Agreement PREDICT Coordinating Center (PCC) and Data Provider (DP) (DHS Form 10036 (12/07)); 6) Memorandum of Agreement - PCC and Data Host (DH)(DHS Form 10037 (12/07)); 7) Authorization Letter for Data Host (DHS Form 10038 (12/07)); 8) Authorization Letter for Data Provider (DHS Form 10039 (12/07)); 9) Sponsorship Letter (DHS Form 10040 (12/07)); 10) Notice of Dataset Access/Application Expiration (DHS Form 10041 (12/07)); 11) Notice for Certificate of Data Destruction (DHS Form 10042 (12/07)). Two new forms are also included – 12) Amendment to Research/User Agreement (10060 (04/10)); 13) Notice of Data Access Expiration (10061 (04/10)).

This notice and request for comments is required by the Paperwork Reduction Act of 1995 (Pub. Law 104-13, 44 U.S.C. chapter 35).

DATES: Comments are encouraged and will be accepted until [INSERT DATE 60-DAYS AFTER DATE OF PUBLICATION IN THE FEDERAL REGISTER].

ADDRESSES: Interested persons are invited to submit comments, identified by docket number **DHS-2010-0043**, by one of the following methods:

- Federal eRulemaking Portal: http://www.regulations.gov. Please follow the instructions for submitting comments.

- E-mail: jeffery.harris@dhs.gov. Please include docket number **DHS-2010-0043** in the subject line of the message.
- Fax: (202) 254-6171. (Not a toll-free number).
- Mail: Science and Technology Directorate, ATTN: OCIO – Jeffery Harris, 245 Murray Drive, Mail Stop 0202, Washington, DC 20528.

FOR FURTHER INFORMATION CONTACT: Jeffery Harris (202) 254-6015 (Not a toll free number).

SUPPLEMENTARY INFORMATION: Interested parties can obtain copies of the Forms Package by calling or writing the point of contact listed above. The content of PREDICT is proprietary datasets that will be used by the Research community in its efforts to build products and technologies that will better protect America's computing infrastructure. Using a secure Web portal, accessible through https://www.predict.org/, the PREDICT Coordinating Center manages a centralized repository that identifies the datasets and their sources and location, and acts as gatekeeper for access and release of the data. All data input to the system is either keyed in by users (Data Providers) or migrated (via upload of XML files).

DHS is particularly interested in comments that:

(1) Evaluate whether the proposed collection of information is necessary for the proper performance of the functions of the agency, including whether the information will have practical utility;

(2) Evaluate the accuracy of the agency's estimate of the burden of the proposed collection of information, including the validity of the methodology and assumptions used;

(3) Suggest ways to enhance the quality, utility, and clarity of the information to be collected; and

(4) Suggest ways to minimize the burden of the collection of information on those who are to respond, including through the use of appropriate automated, electronic, mechanical, or other technological collection techniques or other forms of information technology, e.g., permitting electronic submissions of responses.

The user will complete a portion of the forms online and submit them through the website, while some forms will be printed from the website and faxed to a PREDICT portal administrator. The entire Forms Package will be available on the PREDICT website found at https://www.predict.org.

Overview of this Information Collection:

(1) Type of Information Collection: Information Collection Revision.

(2) Title of the Form/Collection: DHS S&T PREDICT Initiative.

Agency Form Number, if any, and the applicable component of the Department of Homeland Security sponsoring the collection: DHS Science and Technology Directorate, 1) Account Request Form (DHS Form 10029 (12/07)); 2) Request a Dataset Form (DHS Form 10032 (12/07)); 3) My Datasets Form (DHS Form 10033 (12/07)); 4) Memorandum of Agreement - PREDICT (PCC) Coordinating Center and Researcher/User (DHS Form 10035 (12/07)); 5) Memorandum of Agreement PREDICT Coordinating Center (PCC) and Data Provider (DP) (DHS Form 10036 (12/07)); 6) Memorandum of Agreement - PCC and Data Host (DH)(DHS Form 10037 (12/07)); 7) Authorization Letter for Data Host (DHS Form 10038 (12/07)); 8) Authorization Letter for Data Provider (DHS Form 10039 (12/07)); 9) Sponsorship Letter (DHS Form 10040 (12/07)); 10) Notice of Dataset Access/Application Expiration (DHS Form 10041 (12/07)); 11) Notice for Certificate of Data Destruction (DHS Form 10042 (12/07)). Two new forms are also

included – 12) Amendment to Research/User Agreement (10060 (04/10)); 13) Notice of Data Access Expiration (10061 (04/10)).

(3) Affected public who will be asked or required to respond, as well as a brief abstract: Individuals or households, Business or other for-profit, Not-for-profit institutions, Federal government, and State, local, or tribal government; the data gathered will allow the PREDICT initiative to provide a central repository, accessible through a Web-based portal (https://www.predict.org/) that catalogs current computer network operational data, provide secure access to multiple sources of data collected as a result of use and traffic on the Internet, and facilitate data flow among PREDICT participants for the purpose of developing new models, technologies and products that support effective threat assessment and increase cyber security capabilities.

(4) An estimate of the total number of respondents and the amount of time estimated for an average respondent to respond:

a. Estimate of the total number of respondents: 206

b. An estimate of the time for an average respondent to respond: 8 burden hours.

c. An estimate of the total public burden (in hours) associated with the collection: 118 burden hours

Dated: June 3, 2010

Tara O'Toole, M.D., M.P.H.,
Under Secretary for Science and Technology

[FR Doc. 2010-14230 Filed 06/11/2010 at 8:45 am; Publication Date: 06/14/2010]

CRYPTOME

Downloaded 11 June 2010, 09:30 ET

https://www.predict.org/

- PREDICT Overview
- Learn More
- View Dataset Categories
- View Research Publications
- View Research Authors
- Read PREDICT News
- Join the Community
- Login to Portal

Questions?

If you have questions regarding the application process, please review the FAQ.

If you have additional questions or need assistance, please contact the Predict Coordinating Center.

The PREDICT Coordinating Center is currently operated by RTI International in Research Triangle Park, NC.

Fax: (866) 835-0255
Email: PREDICT-Contact@rti.org

PREDICT OVERVIEW

PREDICT, the Protected Repository for the Defense of Infrastructure Against Cyber Threats, is a community of producers of security-relevant network operations data and researchers in networking and information security. Through a centralized repository, it provides developers and evaluators with regularly updated network operations data relevant to cyber defense technology development.

Learn More

Find more information about PREDICT.

View Dataset Categories

Discover the many types of data in the repository. PREDICT has categorized and described datasets from multiple sources. All visitors to the site can view descriptions of the data categories. Only registered users can see the full Data Catalog and apply for access to datasets.

View Research Publications

Read abstracts or download entire papers on the subject of cyber security research. All papers listed use PREDICT data.

View Research Authors

Meet the researchers and find out more about their research.

Read PREDICT News

Find out all the latest happenings with the community and portal.

Join the Community

Identify the PREDICT community role that best meets your needs and become a registered user.

Login to Portal

Registered users only.

ABOUT PREDICT

Background

To help protect America's computing infrastructure and assess threats to it, the Department of Homeland Security (DHS) Science and Technology (S&T) directorate has established a repository for current computer and network operational data. This repository is known as PREDICT, the Protected

Application Review Board

The Application Review Board (ARB) reviews requests for data to evaluate whether the requested datasets are logical for the research described in the request. The board does not judge the merits of the research itself.

Repository for the Defense of Infrastructure Against Cyber Threats.

Protocol

Datasets are made available to qualified cyber defense researchers to help them create and develop new models, technologies, and products to assess cyber threats to the country's computing infrastructure and increase cyber security capabilities. To ensure that business intelligence and individual privacy are not compromised by sharing these datasets, PREDICT has established a data sharing protocol. Key elements of the protocol are:

- Access requirements are established through data sensitivity assessments.
- Access permission is granted after review and approval by independent experts and data provider(s).
- Data usage is subject to legally binding terms and conditions.

PREDICT Coordinating Center

DHS established the PREDICT Coordinating Center (PCC) to centralize data sharing management activities for the distributed repository. The PCC is currently operated by RTI International. Its primary goals are to:

- Provide a central repository, accessible through a web-based portal, that catalogs current computer network and operational data.
- Provide secure access to multiple sources of data collected as a result of use of and traffic on the Internet.
- Facilitate data flow among PREDICT participants for the purpose of developing new models, technologies, and products that support effective threat assessment

Applications for data access may be approved, rejected or approved with conditions. The data provider (see below), who is included in the ARB, has the authority to reject an application for use of its data regardless of the ARB decision.

In addition to the PCC and ARB, roles within PREDICT include:

Researcher

Researchers use the PREDICT data catalog to determine the most applicable dataset for their research and apply for access to the dataset(s). Information on the type of data, time frame for the data snapshot, requirements and restrictions for using the data, as well as criteria for transmission, are provided for each dataset. Multiple datasets may be requested as part of a dataset application. In the case of multiple requests, all requirements for all datasets must be met. Applications will be reviewed by the PREDICT Application Review Board.

Data Provider

Data Providers use the PREDICT portal to upload and register new datasets or retire existing datasets. They specify conditions under which researchers can use their datasets and provide the documents necessary for Researchers to gain approval to use their data. Data Providers also use the portal to monitor who has requested their datasets and communicate their decisions on whether to allow them to be used.

Data Host

Data Hosts use the PREDICT portal to select datasets they wish to host and manage

and increase cyber security capabilities.

The PCC provides the PREDICT web portal as the interface between PREDICT participants, the PCC, and the data repository. The portal's public pages provide general information about PREDICT and access to the account request process. The restricted areas of the portal are accessible only by registered users.

Advisory Board

The Advisory Board advises, assists, consults with, and makes recommendations to the PREDICT Coordinating Center on policy and issues relating to general direction, operation, and scope of the PREDICT project, including the operation of the PREDICT portal.

Members of the Advisory Board are senior in their fields, typically with 20 to 30 years of experience, and are familiar with Internet privacy issues. Members serve a term of one year and meet monthly.

Current board members are:

John McHugh
RedJack, LLC

Catherine Meadows
Naval Research Laboratory

Peter G. Neumann
SRI International

E. Rogers Novak, Jr.
Novak Biddle Venture Partners

Charlotte Scheper
RTI International

information about those datasets. Through the online form a Data Host can supply information to Researchers, such as how the datasets are stored and how they may be transmitted. They can also issue private instructions to the PCC. Access to the portal allows Data Hosts to track expiration dates of datasets they have released to researchers and view other reports on the usage of the data they hold.

Each type of participant is assigned a role in the portal and each role has access permissions applicable to the types of activities that role will perform in the portal.

Data Access Process

To gain access to data, users must be registered users and must submit an application for access to specific datasets. They must also agree to the conditions of use specified for the datasets they request, and they must specify the research they will be conducting using the requested datasets. All applications for datasets are reviewed by the Application Review Board.

Restrictions

Currently, PREDICT datasets are only available to researchers based in the United States.

Privacy Impact Assessment

To learn more about how PREDICT safeguards data, see the US Department of Homeland Security Privacy Impact Assessment for the PREDICT publication.

DATASET CATEGORIES

Category	Description
BGP Routing Table Data	This dataset captures “snapshots” of the topological state of the Internet by archiving Border Gateway Protocol (BGP) routing tables from Internet routers in many locations around the world (these are called Internet Exchange Points). Each routing table expresses the “view” of the Internet from that router’s point in the overall topology and, taken together, all of these views provide a relatively complete roadmap of the connectivity within the Internet Service Provider core of the Internet. This dataset contains only backbone topology information; it does not contain any packet header information or information which relates to individuals. BGP Routing Table Data is used by researchers who study the overall growth patterns of the Internet over time, as well as those who are looking specifically at individual carriers, regions, or resources. It shows historical trends in the utilization of the two principal Internet resources, IP addresses and Autonomous System Numbers (ASN), and this presents the basic backdrop against which many other trends are tracked. Several organizations provide complementary, partly-overlapping, and slightly methodologically different partial views of a large and diverse environment.
BGP Update Messages and BGP Routing Table Dumps	The Boarder Gateway Protocol (BGP) is one of the core routing protocols on today’s Internet and is responsible for exchanging the path information needed to connect autonomous systems (AS) with each other. As one of the critical infrastructure components of the Internet, data from BGP has been used to measure the availability, security, and performance of routing on the Internet including such diverse topics as: the growth of the Internet, the stability of the topology (convergence, flapping), route hijacking, reachability, and AS interconnectedness. This dataset contains full BGP routing table dumps as well as all update messages stored in MRT format (see http://www.ris.ripe.net/source/) as observed by a large tier-2 (regional) ISP over the last six months.
Blackhole Address Space Data	Monitoring packets destined for unused Internet addresses is an important measurement technique in the detection and investigation malicious Internet activity. Since there are no legitimate hosts or devices using the unused addresses, any network traffic observed at those addresses must be the result of misconfiguration, backscatter from spoofed source addresses, or scanning from worms and other network probing. Systems that monitor unused address spaces have a variety of names, including darknets, network telescopes, blackhole monitors, sinkholes, and background radiation monitors. This dataset consists of PCAP formatted (see http://www.tcpdump.org/) packet traces taken at a very large darknet monitor for the last several years. This traffic data can be useful for studying backscatter from distributed denial of service (DDOS) attacks, worm spread (growth rates, population size, and affected population), scanning and backdoor activity, and evaluating various honeypot responders. IP addresses in these files are anonymized by masking the lower order 11-bits.

Category	Description
Full Packet Headers	Full packet headers contain anonymized IP addresses for the sender and recipient. The port numbers identify the application used (e.g., Internet browser or email). No packet content is included in this dataset, although a hash of the data associated with each packet header may be provided. Packet headers can be used for modeling Internet traffic, studying particular attacks, etc.
Internet Topology Data	Internet topology data is created by a program that tries to map the Internet. The program is able to determine which routers are capable of talking to other routers. Internet topology data only shows router connectivity within the Internet core and to external enterprise borders; it does not contain any identifiable information or internal enterprise topology information. This dataset can be used for worm outbreak modeling and simulation, worm containment and countermeasures, zombie distribution for DDOS attacks, vulnerability assessments, longitudinal studies of the evolution of Internet topology and address distribution, Internet topology and address map inference.
Traffic Flows via NetFlow	Netflow is “an embedded instrumentation within IOS Software” used to provide insight into the behavior of networks which has been widely used in applications such as: network traffic accounting, usage-based network billing, network planning, security, Denial of Service monitoring capabilities, and network monitoring. In Netflow, network traffic is represented as flows between two endpoints and records are exported that contain a variety of attributes such as: IP source address, IP destination address, Source port, Destination port, Layer 3 protocol type, packet and byte counts, etc. Several related technologies exist (e.g., jflow, cflowd), but both these and Netflow are being supplanted by an IETF standard (i.e., Internet Protocol Flow Information eXport (IPFIX)). This dataset consists of 1:1 (unsampled) Netflow version 5 collected at the peering edge of a large tier-2 (regional) ISP stored in flow-tools format (see http://code.google.com /p/flow-tools/) for the last six months. IP addresses in these files are anonymized by masking the lower order 11-bits.
VOIP Measurement Data	There are 2200 ISPs registered with data provider that collect data about the quality of Voice Over Internet Protocol (VOIP) connections. This dataset contains only statistical information about VOIP calls. It does not contain phone numbers or content of calls. It consists only of statistical information and a researcher would not be able to identify who was talking, with whom, or about what subject. The dataset is comprised of end-to-end data which characterizes the quality of the paths which VOIP telephone calls take across the global Internet and contains anonymized Session Initiation Protocol (SIP) teardown messages collected from both ends of the conversation. There are three primary pieces of information: IP prefixes of the endpoints of the call, locating it in the Internet topology; autonomous system number of the ISPs at both ends of the connection; country code of Autonomous System Numbers (ASN). All three pieces of information are currently publicly available. In addition, the following data is provided: the call origination time in UTC; the duration of the call in

Category	Description
	seconds; the number of packets delivered in each direction over the course of the call; the average latency of packet transmission in each direction; the loss rate of packets in each direction; the number of packets delivered out-of-order in each direction. This data is very useful to study the quality and security of VOIP calls. It is anticipated that the VOIP statistical data will be used by researchers who wish to compare differential quality of service in similar and dissimilar regions of the Internet, such as across different backbone carriers which utilize different technology or capacity-planning methodologies.

RESEARCH PUBLICATIONS

COTraSE Connection Oriented Traceback in Switched Ethernet

Andreou, M., van Moorsel, A.
Journal of Information Assurance and Security, 2009

Layer 2 Traceback is an important component of end-to-end packet traceback. Whilst IP Traceback identifies the origin network, Layer 2 Traceback extends the process to provide a more fine-grained result. Other known proposals have exposed the difficulties of Layer 2 Traceback in switched ethernet. We build on our earlier "switch-SPIE" and improve in a number of dimensions.

COTraSE Connection Oriented Traceback in Switched Ethernet

Andreou, M., van Moorsel, A.
Journal of Information Assurance and Security, 2009

Layer 2 Traceback is an important component of end-to-end packet traceback. Whilst IP Traceback identifies the origin network, Layer 2 Traceback extends the process to provide a more fine-grained result. Other known proposals have exposed the difficulties of Layer 2 Traceback in switched ethernet. We build on our earlier "switch-SPIE" and improve in a number of dimensions.

On-the-fly statistical classification of internet traffic at application layer based on cluster analysis

Baiocchi, A., Maiolini, G., Molina, G., Rizzi, A.
Proc. Int. Workshop on Computational Intelligence in Security for Information Systems (CISIS), 2008

We address the problem of classifying Internet packet flows according to the application level protocol that generated them.

Using Low-Rate Flow Periodicities for Anomaly Detection: Extended

Bartlett, Genevieve, Heidemann, John, Papadopoulos, Christos
ISI-TR-661, August 5th, 2009

We show that there are several classes of applications that show low-rate periodicity and demonstrate that they are widely deployed on public networks. In this paper we present a new approach to identify changes in low-rate periodic network traffic.

A low complexity visualization tool that helps to perform complex systems analysis

Beiro, M.G., Alvarez-Hamelin, J.I., Busch, J.R.
New Journal of Physics, 2008

In this paper, we present an extension of large network visualization (LaNet-vi), a tool to visualize large scale networks using the k-core decomposition.

The quantitative comparison of computer networks

Brugger, S.T.
THESIS : University of California, Davis, 2009

Being able to compare two traces from computer networks in a quantifiable, holistic, and meaningful way provides us better insight into the network, and enables a range of applications for forensics, administration, application and protocol development, traffic generation, and general network intelligence. In this work we propose a methodology for such a quantifiable comparison of two network traces, be they from different networks, or the same network at different times.

Active Probing to Classify Internet Address Blocks (poster abstract)

Cai, Xue, Heidemann, John
USC/ISI Technical Report ISI-TR-653, August 2008

In this poster we begin to explore the potential of active probing and external classification of address block usage.

Understanding Address Usage in the Visible Internet

Cai, Xue, Heidemann, John
USC/ISI Technical Report ISI-TR-656, February 2009

We provide information about how effectively network addresses blocks appear to be used. We provide new measurements about dynamically managed address space.

Attack diagnosis: throttling distributed denial-of-service attacks close to the attack sources

Chen, R., Park, J.-M.
Proc. 14th Int. Conf. on Computer Communications and Networks (ICCCN), 2005

This paper presents attack diagnosis (AD), a novel attack mitigation scheme that combines the concepts of Pushback and packet marking. AD's architecture is inline with the ideal DDoS attack countermeasure paradigm, in which attack detection is performed near the victim host and attack mitigation is executed close to the attack sources.

Collaborative change detection of DDoS attacks on community and ISP networks

Chen, Y., Hwang, K.

IEEE Symposium on Collaborative Technologies and Systems (CTS), 2006

This paper proposes a collaborative architecture to detect DDoS flooding attacks.

Markov-modulated on/off processes for long-range dependent internet traffic

Clegg, R.
ArXiv Computer Science e-prints, 2006

The aim of this paper is to use a very simple queuing model to compare a number of models from the literature which have been used to replicate the statistical nature of internet traffic and, in particular, the long-range dependence of this traffic.

Towards informative statistical flow inversion

Clegg, R., Haddadi, H., Landa, R., Rio, M.
arXiv:0705.1939v1, 2007

In this paper we propose, implement and test two variants on the sample-and-hold method.

Deployment of an algorithm for large-scale topology discovery

Donnet, B., Raoult, P., Friedman, T., Crovella, M.
IEEE Journal on Selected Areas in Communications, 2006

Topology discovery systems are starting to be introduced in the form of easily and widely deployed software. Unfortunately, the research community has not examined how to perform such measurements efficiently and in a network-friendly manner. This paper describes several contributions towards that end.

Efficient algorithms for large-scale topology discovery

Donnet, B., Raoult, P., Friedman, T., Crovella, M.
Proc. ACM SIGMETRICS Int. Conf. on Measurement and modeling of computer systems, 2005

There is a growing interest in discovery of internet topology at the interface level. A new generation of highly distributed measurement systems is currently being deployed. Unfortunately, the research community has not examined the problem of how to perform such measurements efficiently and in a network-friendly manner. In this paper we make two contributions toward that end.

Improved algorithms for network topology discovery

Donnet, B., Friedman, T., Crovella, M.
Lecture Notes in Computer Science, 2005

We show how to improve the communication scaling properties through the use of Bloom filters to encode a probing stop set.

Retouched Bloom filters: allowing networked applications to trade off selected false positives against false negatives

Donnet, B., Baynat, B., Friedman, T.

Proc. ACM Int. Conf. on Emerging Networking Experiments And Technologies (CoNEXT), 2006

In this paper, we introduce the retouched Bloom filter (RBF), an extension that makes the Bloom filter more flexible by permitting the removal of selected false positives at the expense of generating random false negatives.

Topology discovery using an address prefix based stopping rule

Donnet, B., Friedman, T.
UNICE 2005: Networks and Applications Towards a Ubiquitously Connected World, 2006

Doubletree is a cooperative algorithm that allows the discovery of a large portion of nodes and links in the network while reducing probing redundancy on nodes and destinations as well as the amount of probes sent. In this paper, we propose to reduce more strongly the load on destinations and the communication cost required for the cooperation by introducing a probing stopping rule based on CIDR address prefixes.

Efficient route tracing from a single source

Donnet, B., Raoult, P., Friedman, T.
Computing Research Repository (CoRR), 2006

Traceroute is a networking tool that allows one to discover the path that packets take from a source machine, through the network, to a destination machine. It is not unusual for a route tracing monitor to operate in isolation, rather than from multiple cooperating monitors. Different strategies are required for this case, and this report is the first systematic study of those requirements.

IP Traffic Classification for QoS Guarantees: the Independence of Packets

Dusi, M., Gringoli, F., Salgarelli, L.
Proc. 17th Int. Conf. on Computer Communications and Networks (ICCCN), 2008

In this paper we analyze the impact on flow classification of a hypothesis that is often overlooked, i.e., the tenet that the features of consecutive packets of a given IP flow can be considered statistically independent.

Connectivity measures for internet topologies

Erlebach, T., Moonen, L., Spieksma, F., Vukadinovic, D.
Tech. report, 2005

Four different algorithms have been proposed in the literature for inferring AS relationships using publicly available data from routing tables. We investigate the differences in the graph models produced by these algorithms, focusing on connectivity measures.

Connectivity measures for internet topologies on the level of autonomous systems

Erlebach, T., Moonen, L.S., Spieksma, F.C.R., Vukadinovic, D.
Operations Research, 2009

We consider the adaptation of the classical connectivity measures to the valley-free path model, where it is -hard to compute them.

Bitmap algorithms for counting active flows on high speed links

Estan, C., Varghese, G., Fisk, M.
IEEE/ACM Transactions on Networking (TON), 2006

This paper presents a family of bitmap algorithms that address the problem of counting the number of distinct header patterns (flows) seen on a high-speed link.

New directions in traffic measurement and accounting: focusing on the elephants, ignoring the mice

Estan, C., Varghese, G.
ACM Transactions on Computer Systems (TOCS), 2003

Accurate network traffic measurement is required for accounting, bandwidth provisioning and detecting DoS attacks. These applications see the traffic as a collection of flows they need to measure. We propose two novel and scalable algorithms for identifying the large flows: sample and hold and multistage filters, which take a constant number of memory references per packet and use a small amount of memory.

Support Vector Machines for TCP traffic classification

Este, A., Gringoli, F., Salgarelli, L.
Computer Networks, 2009

In this paper we describe an approach to traffic classification based on Support Vector Machines (SVM). We apply one of the approaches to solving multi-class problems with SVMs to the task of statistical traffic classification, and describe a simple optimization algorithm that allows the classifier to perform correctly with as little training as a few hundred samples.

Protecting TCP services from denial of service attacks

Farhat, H.
Proc. 2006 SIGCOMM workshop on Large-scale attack defense, 2006

In this paper, we present a scheme that protects legitimate traffic from the large volume of attackers packets during a DDoS attack.

The inframetric model for the internet

Fraigniaud, P., Febhar, E., Viennot, F.
Proc. IEEE 27th Conf. on Computer Communications (INFOCOM), 2008

A large amount of algorithms have recently been designed for the Internet under the assumption that the distance defined by the round-trip delay (RTT) is a metric. Many of these algorithms rely on the assumption that the metric has bounded ball growth or bounded doubling dimension. This paper analyzes the validity of these assumptions and proposes a tractable model matching experimental observations.

Relevance of massively distributed explorations of the internet topology: qualitative results

Guillaume, J.-L., Latapy, M., Magoni, D.
Computer Networks, 2006

Internet maps are generally constructed using the traceroute tool from a few sources to many destinations. It appeared recently that this exploration process gives a partial and biased view of the real topology, which leads to the idea of increasing the number of sources to improve the quality of the maps. In this paper, we present a set of experiments we have conducted to evaluate the relevance of this approach.

Analyzing router responsiveness to measurement probes

Gunes, M., Sarac, K.
Lecture Notes in Computer Science (LNCS), 2009

In this paper, we conduct an experimental study to understand the responsiveness of routers to active probing both from a historical perspective and current practices.

Revisiting the issues on netflow sample and export performance

Haddadi, H., Landa, R., Rio, M., Bhatti, S.
http://arxiv.org, 2006

In this paper, we assess the performance of the sampling process as used in NetFlow in detail, and we discuss some techniques for the compensation of loss of monitoring detail.

Uses and Challenges for Network Datasets

Heidemann, John USC/ISI
March 2009

The goal of this paper is to explore the research questions facing the Internet today, the datasets needed to answer those questions, and the challenges to using those datasets. We suggest several practices that have proven important in use of current datasets, and open challenges to improve use of network data.

A histogram-based stochastic process for finite buffer occupancy analysis

Hernandez-Orallo, E., Vila-Carbo, J.
Proc. Int. Conf. on Performance Evaluation Methodologies and Tools, 2007

This paper proposes to use histograms for characterising network traffic and a simple stochastic process for network performance analysis.

Network performance analysis based on histogram workload models

Hernandez-Orallo, E., Vila-Carbo, J.
Proc. IEEE/ACM Int. Symp. on "Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS), 2007

Network performance analysis relies mainly on two models: a workload model and a performance model. This paper proposes to use histograms for characterising the arrival workloads and a

performance model based on a stochastic process.

Entropy based flow aggregation

Hu, Y., Chiu, D.-M., Lui, J.
Networking 2006, 2006

In this paper, we propose an entropy based flow aggregation algorithm, which not only alleviates the problem in memory and export bandwidth, but also maximizes the accuracy of legitimate flows.

Network traffic analysis using traffic dispersion graphs (TDGs): techniques and hardware implementation

Iliofotou, M., Pappu, P., Faloutsos, M., Mitzenmacher, M., Singh, S., Varghese, G.
Technical Report UCR-CS-2007-05001, 2007

In this work, we propose the use of Traffic Dispersion Graphs (TDGs) as a powerful way to monitor, analyze, and visualize network traffic.

On properties of multicast routing trees

Janic, M., Van Mieghem, P.
International Journal of Communication Systems, 2006

In this paper we propose analytical expressions that could be used to estimate the gain of network-layer multicast. We examine the reliability of traceroute data and of traceroutes-based conclusions.

Why is the internet traffic bursty in short (sub-RTT) timescales?

Jiang, H., Dovrolis, C.
ACM SIGMETRICS Performance Evaluation Review, 2006

Internet traffic exhibits multifaceted burstiness and correlation structure over a wide span of time scales. Previous work analyzed this structure in terms of heavy-tailed session characteristics, as well as TCP timeouts and congestion avoidance, in relatively long time scales. We focus on shorter scales, typically less than 100-1000 milliseconds. Our objective is to identify the actual mechanisms that are responsible for creating bursty traffic in those scales.

Small-world characteristics of internet topologies and implications on multicast scaling

Jin, S., Bestavros, A.
Computer Networks, 2006

This paper tries to understand how small-world behavior arises in the Internet topologies and how it impacts the performance of multicast techniques.

Long-range dependence: ten years of Internet traffic modeling

Karagiannis, T., Molle, M., Faloutsos, M.
IEEE Internet Computing, 2004

The authors outline long-range dependence (LRD) findings in network traffic and explore the

current lack of accuracy and robustness in LRD estimation.

The power of slicing in internet flow measurement

Kompella, R., Estan, C.
Proc. 5th ACM SIGCOMM Conf. on Internet Measurement (IMC), 2005

In this paper, we propose Flow Slices, a solution inspired from previous enhancements to NetFlow such as Smart Sampling, Adaptive NetFlow (ANF). Flow Slices, in contrast to NetFlow, controls the three resource bottlenecks at the router using separate "tuning knobs."

Lobby index in networks

Korn, A., Schubert, A., Telcs, A.
Physica A: Statistical Mechanics and its Applications, 2009

We propose a new node centrality measure in networks, the lobby index, which is inspired by Hirsch's h-index.

Exploiting locality to ameliorate packet queue contention and serialization

Kumar, S., Maschmeyer, J., Crowley, P.
Proc. 3rd ACM Int. Conf. on Computing Frontiers, 2006

In this paper we observe that the worst-case scenario for packet queuing coincides with the best-case scenario for caches: i.e., when locality exists and the majority of packets are destined for a small number of queues. The main contribution of this work is the queuing cache, which consists of a hardware cache and a closely coupled queuing engine that implements queue operations.

Efficient simulation of large-scale P2P networks: modeling network transmission times

Kunzmann, G., Nagel, R., Hossfeld, T., Binzenhofer, A., Eger, K.
EUROMICRO 15th Int. Conf. on Parallel, Distributed and Network-Based Processing (PDP), 2004

In this paper we take a closer look at the network layer. We compare the most commonly-used network models and present a very efficient model for applying real-world network transmission times in large scale simulations.

Complex network measurements: estimating the relevance of observed properties

Latapy, M., Magnien, C.
Proc. 27th IEEE Conf. on Computer Communications (INFOCOM), 2008

Topological information on complex networks is only available through intricate measurement procedures. Until recently, most studies assumed these procedures lead to samples large enough to be representative of the whole. Recent contributions proved this assumption may be misleading. We provide here the first practical methodology to distinguish between cases where it is misleading, and cases where the observed properties may be trusted.

Measuring fundamental properties of real-world complex networks

Latapy, M., Magnien, C.

Computing Research Repository (CoRR), 2006

Data on complex networks is only available through intricate measurement procedures. Until recently, most studies assumed these procedures lead to samples large enough to be representative of the whole. Recent contributions proved that this approach may be misleading. We provide here the first practical way to distinguish between cases where it is misleading, and cases where the observed properties may be trusted.

Describing and simulating internet routes

Leguay, J., Latapy, M., Friedman, T., Salamatian, K.
Computer Networks, 2007

This contribution deals with actual routes followed by packets in the Internet at the IP level. We propose a set of statistical properties to analyze such routes. We use the results to suggest and evaluate methods for generating artificial routes suitable for simulation purposes.

A hardware packet re-sequencer unit for network processors

Meitinger, M., Ohlendorf, R., Wild, T., Herkersdorf, A.
Architecture of Computing Systems (ARCS), 2008

In this paper, we describe a Hardware Re-Sequencer Unit for Network Processors.

An application-aware load balancing strategy for network processors

Ohlendorf, R., Meitinger, M., Wild, T., A. Herkersdorf, A.
Proc. 5th Int. Conf. on High performance embedded architectures and compilers (HIPEAC 2010), 2009

This paper presents and compares different load balancing strategies in multi-core network processor (NP) chips.

FlexPath NP-flexible, dynamically reconfigurable processing paths in network processors

Ohlendorf, R., Meitinger, M., Wild, T., Herkersdorf, A.
Dynamically reconfigurable systems, architectures, design methods and applications, 2010

We propose to extend state-of-the-art processor-centric NP architectures with specific hardware units in order to classify the incoming traffic into separate processing classes. We propose to offload significant shares of the traffic to a dedicated hardware path in order to bypass the CPU cluster and save precious programmable processing resources.

Linguistic summarization of network traffic flows

Pouzols, F.M., Barriga, A., Lopez, D.R., Sanchez-Solano, S.
Proc. IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE), 2008

We address, by means of fuzzy linguistic summaries, two related problems: summarizing network flow statistics and making these statistics human-readable. Two complementary summarization methods are developed.

Confidence-weighted marginal utility analyses for improved internet mapping

Prince, C., Wyatt, D.
Tech. report, Class Project, 2004

Many techniques have been proposed for discerning internet topology, but not so many of them have been studied for their utility: the returns in information that they provide as their number of measurements increases. We define a probabilistic model for representing a map built on uncertain observations. We propose new utility metrics and reuse existing metrics in combination with this model to examine the marginal utility of observations for maps made with different tools.

Joint entropy analysis model for DDoS attack detection

Rahmani, H., Sahli, N., Kamoun, F.
Proc. 5th Int. Conf. on Information Assurance and Security (IAS), 2009

While previous work has demonstrated the benefits of entropy-based anomaly detection, there has been little effort to understand the detection power of using joint entropy analysis of multiple traffic distributions. We observe that the time series of IP-flow number and aggregate traffic size are strongly statistically dependant.

On the correlation of geographic and network proximity at internet edges and its implications for mobile unicast and multicast routing

Schmidt, T., Wählisch, M., Zhang, Y.
Proc. 6th Int. Conf. on Networking (ICN), 2007

As continuous mobility handovers necessarily occur between access routers located in geographic vicinity, this paper investigates on the hypothesis that geographically adjacent edge networks attain a reduced network distances as compared to arbitrary Internet nodes.

A proactive test based differentiation to mitigate low rate DoS attacks

Shevtekar, A., N. Ansari, N.
Proc. 16th IEEE Int. Conf. on Computer Communications and Networks (ICCCN), 2007

Low rate DoS attacks are emerging threats to the TCP traffic, and the VoIP traffic in the Internet. They are hard to detect as they intelligently send attack traffic inside the network to evade current router based congestion control mechanisms. We propose a practical attack model in which botnets that can pose a serious threat to the Internet are considered.

A router-based technique to mitigate reduction of quality (RoQ) attacks

Shevtekar, A., Ansari, N.
Computer Networks, 2008

We propose a router-based technique to mitigate the stealthy reduction of quality (RoQ) attacks at the routers in the Internet.

Geometric exploration of the landmark selection problem

Tang, L., Crovella, M.

Lecture Notes in Computer Science, 2004

Internet coordinate systems appear promising as a method for estimating network distance without direct measurement and allowing scalable configuration of emerging applications. All such systems rely on landmarks. In this paper we explore fast algorithms for landmark selection.

Levy flights and fractal modeling of internet traffic

Terdik, G., Gyires, T.
IEEE/ACM Transactions on Networking, 2009

The major contribution of the paper is the application of two new analytical methods. We apply the theory of smoothly truncated Levy flights and the linear fractal model in examining the variability of Internet traffic from self-similar to Poisson.

Detection of Low-Rate Attacks in Computer Networks

Thatte, Gautam, Mitra, Urbashi, Heidemann, John
April 2008

This paper develops two parametric methods to detect low-rate denial-of-service attacks and other similar near-periodic traffic, without the need for flow separation.

Parametric Methods for Anomaly Detection in Aggregate Traffic

Thatte, Gautam, Mitra, Urbashi, Heidemann, John
USC/ISI TECHNICAL REPORT ISI-TR-663, August 2009

This paper develops parametric methods to detect network anomalies using only aggregate traffic statistics in contrast to other works requiring flow separation, even when the anomaly is a small fraction of the total traffic.

Detecting evasion attacks at high speeds without reassembly

Varghese, G., Fingerhut, J., Bonomi, F.
Proc. Conf on Applications, technologies, architectures, and protocols for computer communications, 2006

Both the state and processing requirements of reassembly and normalization are barriers to scalability for an IPS at speeds higher than 10 Gbps. In this paper, we suggest breaking with this paradigm using an approach we call Split-Detect.

Multifractality in TCP/IP traffic: the case against

Veitch, D., Hohn, N., Abry, P.
Computer Networks, 2005

In this paper we review the evidence for multifractal behaviour of aggregate TCP traffic, and show that in many ways it is weak. Our study includes classic traces and very recent ones.

Realistic and responsive network traffic generation

Vishwanath, K., Vahdat, A.

Proc. ACM SIGCOMM Conf. on Applications, technologies, architectures, and protocols for computer communications, 2006

This paper presents Swing, a closed-loop, network-responsive traffic generator that accurately captures the packet interactions of a range of applications using a simple structural model.

Easily-implemented adaptive packet sampling for high speed networks flow measurement

Wang, H., Lin, Y., Jin, Y., Cheng, S.
Proc. Conf Computational Science (ICCS), 2006

An Easily-implemented Adaptive Packet Sampling (EAPS) is presented in this paper, which overcomes the shortcoming of NetFlow and Adaptive Netflow.

A TCP connection establishment filter: symmetric connection detection

Whitehead, B., Lung, C.-H., Rabinovitch, P.
Proc. IEEE Int. Conf. on Communications (ICC), 2007

Symmetric connection detection (SCD) is a method of filtering TCP sessions, passing only those sessions which become fully established. SCD can benefit network monitoring applications that are only interested fully established TCP connections by reducing processing requirements.

Mark-aided distributed filtering by using neural network for DDoS defense

Xiang, Y., Zhou, W.
Proc. IEEE Global Telecommunications Conference (GLOBECOM), 2005

The immediate task of DDoS defense is to provide as much resources as possible to legitimate users when there is an attack. Unfortunately most current defense approaches can not efficiently detect and filter out attack traffic. Our approach is to find the network anomalies by using neural network, deploy the system at distributed routers, identify the attack packets, and then filter them.

Modeling the IPv6 internet AS-level topology

Xiao, B., Liu, L., Guo, X., Xu, K.
Physica A: Statistical Mechanics and its Applications, 2009

To measure the IPv6 internet AS-level topology, a network topology discovery system, called Dolphin, was developed. We propose a new model based on the two major factors affecting the exponent of the EBA model. We demonstrate how this model can be successfully used to reproduce the topology of the IPv6 Internet.

An improved BA model for router-level internet macroscopic topology

Xu, Y., Zhao, H.
IAENG Int. Journal of Computer Science, 2009

Router-level Internet macroscopic topology modeling is studied in this paper.

FIT: Fast Internet Traceback

Yaar, A., Perrig, A., Song, D.
Proc. 24th Annual Joint Conf. of the IEEE Computer and Communications Societies (INFOCOM), 2005

Traceback mechanisms are a critical part of the defense against IP spoofing and DoS attacks, as well as being of forensic value to law enforcement. Currently proposed IP traceback mechanisms are inadequate to address the traceback problem. We propose fast Internet traceback (FIT), a new packet marking approach that significantly improves IP traceback in several dimensions.

Pi: a path identification mechanism to defend against DDoS attacks

Yaar, A., Perrig, A., Song, D.
IEEE Symposium on Security and Privacy, 2003

Distributed denial of service attacks continue to plague the Internet. Defense against these attacks is complicated by spoofed source IP addresses, which make it difficult to determine a packet's true origin. We propose Pi (short for path identifier), a new packet marking approach in which a path fingerprint is embedded in each packet, enabling a victim to identify packets traversing the same paths through the Internet on a per packet basis, regardless of source IP address spoofing.

SIFF: a Stateless Internet Flow Filter to mitigate DDoS flooding attacks

Yaar, A., Perrig, A., Song, D.
IEEE Symposium on Security and Privacy, 2004

In this paper, we present SIFF, a Stateless Internet Flow Filter, which allows an end-host to selectively stop individual flows from reaching its network, without any common assumptions.

StackPi: new packet marking and filtering mechanisms for DDoS and IP spoofing defense

Yaar, A., Perrig, A., Song, D.
IEEE Journal on Selected Areas in Communications, 2006

In this paper, we propose the StackPi marking, a new packet marking scheme based on Pi, and new filtering mechanisms. The StackPi marking scheme consists of two new marking methods that substantially improve Pi's incremental deployment performance: Stack-based marking and write-ahead marking.

ProgME: towards programmable network measurement

Yuan, L., Chuah, C., Mohapatra, P.
ACM SIGCOMM Computer Communication Review, 2007

We present ProgME, a Programmable Measurement architecture based on a novel concept of flowset - arbitrary set of flows defined according to application requirements and/or traffic conditions. Through a simple flowset composition language, ProgME can incorporate application requirements, adapt itself to circumvent the challenges on scalability posed by the large number of flows, and achieve a better application-perceived accuracy.

TOPO: a topology-aware single packet attack traceback scheme

Zhang, L., Guan, Y.
Securecomm and Workshops, 2006

To reduce the impact of unavoidable collisions in Bloom filters, we propose a topology-aware single packet IP traceback system, namely TOPO.

RESEARCH AUTHORS

Patrice Abry

Ecole Normale Supérieure de Lyon, Laboratoire de Physique
J.I. Alvarez-Hamelin

Universidad de Buenos Aires, Facultad de Ingeniería
Marios Andreou

Newcastle University, School of Computing Science
Nirwan Ansari

New Jersey Institute of Technology, Department of Electrical and Computer Engineering
Andrea Baiocchi

University of Roma, INFOCOM
A. Barriga

Instituto de Microelectrónica de Sevilla
Genevieve Bartlett

University of Southern California, Information Sciences Institute
Bruno Baynat

Universite Pierre et Marie Curie
M.G. Beiro

Universidad de Buenos Aires, Facultad de Ingeniería
Azer Bestavros

Boston University, Computer Science Department
Saleem Bhatti

University of St. Andrews, School of Computer Science

Andreas Binzenhofer

Tech. Univ. of Munich, Institute for Communication Networks
Flavio Bonomi

Cisco Systems
Terry Brugger

University of California, Davis
J.R. Busch

Universidad de Buenos Aires, Facultad de Ingeniería
Xue Cai

University of Southern California, Department of Computer Science
R. Chen

Bradley Department of Electrical & Computing Engineering, Virginia
Yu Chen

University of Southern California, Los Angeles
Shiduan Cheng

University of Posts and Telecommunications
Dah-Ming Chiu

The Chinese University of Hong Kong
Chee-Nee Chuah

University of California, Davis
Richard Clegg

University of York, Department of Mathematics
Mark Crovella

Boston University, Department of Computer Science
Patrick Crowley

Washington University
Benoit Donnet

Laboratoire LiP6-CNRS, Université Pierre & Marie Curie
Constantinos Dovrolis

Georgia Tech

M. Dusi

Universita' degli Studi di Brescia
Kolja Eger

Technical University of Munich
Thomas Erlebach

University of Leicester, Department of Computer Science
Cristian Estan

University of California, San Diego
Alice Este

Università degli Studi di Brescia
Michalis Faloutsos

California University, Riverside, Dept. of Computer Science & Engineering
Hikmat Farhat

Notre Dame University, Lebanon
E. Febhar

Universite Paris, CNRS
J. Andrew Fingerhut

Washington University
M. Fisk

Wisconsin University, Department of Computer Science
P. Fraigniaud

Universite Paris, CNRS
Timur Friedman

Université Pierre & Marie Curie, CNRS
Francesco Gringoli

Universita' Degli Studi di Brescia, DEA
Yong Guan

Iowa State University, Department of Electrical and Computer Engineering
Jean-Loup Guillaume

Université Paris, CNRS

Mehmet Gunes

University of Nevada
Xiao-chen Guo

Beihang University, School of Computer Science and Engineering
T. Gyires

Univ. of Debrecen, Faculty of Computer Science
Hamad Haddadi

University College, Department of Electronic and Electrical Engineering
John Heidemann

University of Southern California, Information Sciences Institute
Andreas Herkersdorf

University of Technology Munich, Institute for Integrated Systems
Enrique Hernandez-Orallo

Polytechnic University of Valencia, Dept. de Inf. de Sist. y Comput.
Nicolas Hohn

University of Melbourne, Department of Electrical and Electronic Engineering
Tobias Hossfeld

University of Technology Munich, Institute for Communication Networks
Yan Hu

The Chinese University of Hong Kong
Kai Hwang

University of Southern California
Marios Iliofotou

California University, Riverside, Department of Computer Science & Engineering
Milena Janic

Delft University of Technology, Faculty of EEMCS
Hao Jiang

Georgia Tech
Shudong Jin

Case Western Reserve University, Department of Electrical Engineering and Computer Science

Yuehui Jin

State Key Laboratory of Networking and Switching Technology
F. Kamoun

National School for Computer Science of Tunis, CRISTAL Lab
T. Karagiannis

California University, Riverside, Department of Computer Science & Engineering
Ramana Rao Kompella

University of Southern California, San Diego
A. Korn

University of Technology and Economics, Department of Telecommunications and Media Informatics
Sailesh Kumar

Washington University
Gerald Kunzmann

University of Technology Munich, Institute for Communication Networks
Raul Landa

University College London, Department of Electronic and Electrical Engineering
Matthieu Latapy

University Pierre & Marie Curie, CNRS
Jeremie Leguay

University Pierre & Marie Curie, CNRS
Yu Lin

State Key Laboratory of Networking and Switching Technology
Lian-dong Liu

Beihang University, School of Computer Science and Engineering
D.R. Lopez

Instituto de Microelectrónica de Sevilla, Consejo Super. de Investig. Cientificas
John Lui

The Chinese University of Hong Kong
C.-H. Lung

Carleton University

Clemence Magnien

University Pierre & Marie Curie, CNRS
Damien Magoni

Université Strasbourg
Gianluca Maiolini

University of Roma, INFOCOM
John Maschmeyer

Washington University
Michael Meitinger

University of Technology Munich, Institute for Integrated Systems
Urbashi Mitra

University of Southern California, Department of Electrical Engineering
Michael Mitzenmacher

Harvard University
Prasant Mohapatra

University of California, Davis
Giacomo Molina

University of Roma, INFOCOM
M. Molle

California University, Riverside, Department of Computer Science & Engineering
Linda Moonen

Katholieke Universiteit Leuven, Operations Research Group
Robert Nagel

Tech. Univ. of Munich, Inst. of Commun. Networks
Rainer Ohlendorf

University of Technology Munich, Institute for Integrated Systems
Christos Papadopoulos

Colorado State University, Department of Computer Science
Prashanth Pappu

Cisco

J.-M. Park

Polytechnic Institute & State University
A. Perrig

Carnegie Mellon University
F.M. Pouzols

Instituto de Microelectrónica de Sevilla, Consejo Super. de Investig. Cientificas
Craig Prince

University of Washington
P. Rabinovitch

Carleton University
H. Rahmani

National School for Computer Science of Tunis, CRISTAL Lab
Philippe Raoult

Universite' Pierre et Marie Curie
Miguel Rio

University College London, Department of Electronic and Electrical Engineering
Antonello Rizzi

University of Roma, INFOCOM
N. Sahli

National School for Computer Science of Tunis, CRISTAL Lab
Kave Salamatian

Université Paris, CNRS
Luca Salgarelli

Univ. degli Studi di Brescia, DEA
S. Sanchez-Solano

Instituto de Microelectrónica de Sevilla, Consejo Super. de Investig. Cientificas
Kamil Sarac

University of Texas at Dallas
T. Schmidt

HAW Hamburg

A. Schubert

Institute for Research Organisation
Amey Shevtekar

New Jersey Institute of Technology, Department of Electrical and Computer Engineering
Sumeet Singh

Cisco
D. Song

Carnegie Mellon University
Frits Spieksma

Katholieke Universiteit, Operations Research Group
Liying Tang

Boston University, Department of Computer Science
A. Telcs

University of Technology and Economics Budapest, Department of Computer Science and Information Theory
G. Terdik

University of Debrecen, Faculty of Computer Science
Gautam Thatte

University of Southern California, Department of Electrical Engineering
Amin Vahdat

University of California, San Diego
Piet Van Mieghem

Delft University of Technology, Faculty of EEMCS
Aad van Moorsel

Newcastle University, School of Computing Science
George Varghese

University of California, San Diego, Dept. of Comput. Sci.
Darryl Veitch

University of Melbourne, Department of Electrical and Electronic Engineering
F. Viennot

University Paris, CNRS

Joan Vila-Carbo

Universitat Politècnica de València, Dept. de Inf. de Sist. y Comput.
Kashi Vishwanath

University of California, San Diego
Danica Vukadinovic

Computer Engineering and Networks Laboratory (TIK)
M. Wählisch

HAW Hamburg
Hongbo Wang

State Key Laboratory of Networking and Switching Technology
B. Whitehead

Carleton University
Thomas Wild

Technische Universität München
Danny Wyatt

University of Washington
Yang Xiang

Deakin University, School of Information Technology
Bo Xiao

Beihang University, School of Computer Science and Engineering
Ke Xu

Beihang University, School of Computer Science and Engineering
Ye Xu

Shenyang Ligong University
A. Yaar

Carnegie Mellon University
Lihua Yuan

University of California, Davis
Linfeng Zhang

Iowa State University, Department of Computer Science

Y. Zhang

HAW Hamburg

Hai Zhao

Shenyang Ligong University

Wanlei Zhou

Deakin University, School of Information Technology

PREDICT NEWS

Workshop Series: Issues in Network Security Research

An ongoing series of five workshops (four in 2009, one in March 2010, with more planned) is focused on an investigation of ethical issues underlying research in network security. A document, The Menlo Report, is being prepared.

PREDICT Privacy Workshop

A workshop was held 16-17 February 2010 in Menlo Park, California, on the various forms of anonymity that could be useful in enhancing data privacy in PREDICT.

JOIN THE COMMUNITY

Account Request Process

To apply for an account:

- Complete the form and click Submit.
- The PCC will review your account and email any necessary additional documents to you.
  Complete any documents emailed to you and have them signed by someone in your organization with signature authority.
- Fax the completed and signed document(s) to the PCC.

Required fields are marked with *

Select Your Role	Enter Your Temporary Password
Selecting more than one role is permitted. ☐ Researcher Sponsorship Letter Example	As a means to authenticate you once your account is approved, please enter a password below. This will be the password that you use once your account request is

approved. Your password must conform to the PREDICT password standards.

Researcher MOA Example

Password*

☐ Data Host

☐ Data Provider

Re-Enter Password*

Your Contact Details

First Name*

Last Name*

Street 1*

Street 2

City*

State*

Zip Code*

Office Phone*

Home Phone

Cell Phone

Fax

EMail*

Your Sponsoring Organization

Organization Details

Org. Name*

Street 1*

Street 2

City*

State*

Zip Code*

Authorized Representative Details

First Name*

Last Name*

Phone*

EMail*

☐ I would like to receive email updates when new datasets are added or updated to the catalog

Our Rules Of Behavior

All users applying for accounts on PREDICT must acknowledge and agree to abide by the PREDICT Rules of Behavior. Check the box below to acknowledge that you have read the rules and agree to follow them. You will not be able to submit this form unless you check the box.

☐ I agree to abide by the PREDICT Rules of Behavior

Submit Account Request

 Privacy Act Notice

DHS Authority to Collect This Information: The Homeland Security Act of 2002 [Public Law 107-296, §302(4)] authorizes the Science and Technology Directorate to conduct 'basic and applied research, development, demonstration, testing, and evaluation activities that are relevant to any or all elements of the Department, through both intramural and extramural programs.' In exercising its responsibility under the Homeland Security Act, S&T is authorized to collect information, as appropriate, to support R&D related to improving the security of the homeland.

Principal Purpose: DHS collects name, organization and title (if any), email address, home and/or work address, and telephone numbers for the purpose of contacting individuals regarding the PREDICT project and/or their involvement with PREDICT. When using the PREDICT portal, your IP address, user name, browser type, and access times will be collected for the purpose of conducting research about your opinion of current services or of potential new services that may be offered, or facilitating the operation of the PREDICT service, or maintaining quality of the service, or providing general statistics regarding use of the PREDICT Web site.

Routine Uses and Sharing: Some of your information will be disclosed to PREDICT team members, such as data hosts, data providers, PREDICT contractors, the Predict Coordinating Center, the advisory board, and review board members to help us deliver requested PREDICT services and operate the PREDICT Web site and deliver the services you have requested. Unless you consent otherwise, this information will not be used for any purpose other than those stated above. However, DHS may release this information for an individual on a case-by-case basis as described in the DHS/ALL-002 System of Records Notice (SORN), which can be found at: www.dhs.gov/privacy.

Disclosure: Furnishing this information is entirely voluntary; however, failure to furnish at least the minimum information required to register (to include full name and email address) will prevent you from obtaining authorization to access system.

PRA Burden Statement: An agency may not conduct or sponsor an information collection and a person is not required to respond to this information collection unless it displays a current valid OMB control number and an expiration date. The control number for this collection is 1640-0012 and this form will expire on 8/31/2010. The estimated average time to complete this form is 15 minutes per respondent. If you have any comments regarding the burden estimate you can write to Department of Homeland Security, Science and Technology Directorate, Washington, DC 20528.

DHS Form 10029 (12/07)

PREDICT
SPONSORSHIP LETTER FOR PREDICT ACCOUNT

READ THESE INSTRUCTIONS CAREFULLY BEFORE PROCEEDING

Thank you for your interest in joining the PREDICT community. All Researchers must have a sponsoring organization in order to obtain a PREDICT account. A PREDICT account enables a Researcher to access the PREDICT catalog of datasets and to request the use of those datasets. A completed and signed sponsorship letter must be received by the PREDICT Coordinating Center before your application for an account as a Researcher can be considered. Sponsorship letters must be signed by a supervisor or manager with authority to act on behalf of your organization.

Definition of Researcher

A Researcher may be an individual or it may be an entity, such as a corporation that desires to have a team of personnel conduct specific cyber security research and development (R&D). If the Researcher is an entity, the entity must name a Data Custodian who is the person designated to have a PREDICT account and request datasets on behalf of that entity. Entities may have more than one Data Custodian, with each person having a PREDICT account. Researchers in an academic environment usually have individual PREDICT accounts and are sponsored as individuals by their institution. An individual Researcher may involve others in the R&D project that he/she plans to conduct using PREDICT datasets.

INSTRUCTIONS

Researchers seeking an individual PREDICT account must submit the attached Sponsorship Letter *on the sponsoring organization's letterhead* to the PREDICT Coordinating Center (PCC). Check the box: Sponsorship of Individual as Researcher. The letter must be signed by a supervisor or other appropriate manager from the sponsoring institution who has authority to act on behalf of the organization. Individuals named as Researchers must be employed by or affiliated with the Sponsoring Institution. The completed and signed Sponsorship Letter must be received by the PREDICT Coordinating Center (PCC) before a PREDICT account will be assigned. After the Sponsorship Letter has been accepted by the PCC, the actual application for a PREDICT account is made by the individual Researcher(s) through the PREDICT portal at http://www.predict.org.

Organizations who are seeking to be an entity acting as a Researcher must submit the attached Sponsorship Letter *on the sponsoring organization's letterhead* to the PREDICT Coordinating Center (PCC). Check the box: Sponsorship of Entity as Researcher. The letter must be signed by a supervisor or person who has authority to act on behalf of the organization. The Sponsorship Letter must designate a Data Custodian(s) to have a PREDICT account and manage research for the organization. Reminder: a Data Custodian is the individual with primary

DHS Authority to Collect This Information: The Homeland Security Act of 2002 [Public Law 107-296, §302(4)] authorizes the Science and Technology Directorate to conduct "basic and applied research, development, demonstration, testing, and evaluation activities that are relevant to any or all elements of the Department, through both intramural and extramural programs." In exercising its responsibility under the Homeland Security Act, S&T is authorized to collect information, as appropriate, to support R&D related to improving the security of the homeland. **Principal Purpose:** DHS collects name, organization and title (if any), email address, home and/or work address, and telephone numbers for the purpose of contacting individuals regarding the PREDICT project and/or their involvement with PREDICT. **Routine Uses and Sharing:** Some of your information will be disclosed to PREDICT team members, such as data hosts, data providers, PREDICT contractors, the Predict Coordinating Center, the advisory board, and review board members to help us deliver requested PREDICT services and operate the PREDICT Web site and deliver the services you have requested. Unless you consent otherwise, this information will not be used for any purpose other than those stated above. However, DHS may release this information for an individual on a case-by-case basis as described in the DHS/ALL-002 System of Records Notice (SORN), which can be found at: www.dhs.gov/privacy. **Disclosure:** Furnishing this information is entirely voluntary; however, failure to furnish at least the minimum information required to register (to include full name and email address,) will prevent you from obtaining authorization to access system.

PRA Burden Statement: A U.S. Government agency may not conduct or sponsor an information collection and a person is not required to respond to this information collection unless it displays a current valid OMB control number and an expiration date. The control number for this collection is 1640-0012 and this form will expire on 08/31/2010. The estimated average time to complete this form is 60 minutes per respondent. If you have any comments regarding the burden estimate you can write to Department of Homeland Security, Science and Technology Directorate, Washington, DC 20528.

responsibility for the receipt, security, oversight, use, and return of the Data that is obtained from PREDICT for a particular research effort. The Data Custodian does not have to be an employee of or affiliated with the Sponsoring Institution. After the Sponsorship Letter has been accepted by the PCC, the Data Custodian(s) apply for PREDICT accounts through the PREDICT portal at http://www.predict.org.

INSTRUCTIONS FOR ALL APPLICANTS

- Put the text of the Sponsorship letter onto your organization's letterhead.
- Fill in appropriate names, dates, and other requested information; do not omit any of the information, as incomplete letters will be returned and the process delayed accordingly.
- Do not use abbreviations for organization names, schools, departments, etc.
- Fax the signed (866) 835-0255 (toll free) or email a PDF file of the letter to PREDICT-contact@rti.org.
- PCC will notify you of acceptance/rejection of the Sponsorship Letter usually within one week from receipt.

QUESTIONS OR NEED ASSISTANCE?

Contact the PCC for assistance via email at PREDICT-contact@rti.org.

_____
Date

RTI International
PREDICT Coordinating Center
Attn: Renee Karlsen
P.O. Box 12194
Research Triangle Park, NC 27709-2194

SUBJECT: Sponsorship Letter for PREDICT Account

Dear PREDICT Coordinating Center:

I am sending this Sponsorship Letter for Researcher access to PREDICT data by _____ (enter name of entity or individual(s) being sponsored). I understand that a Sponsorship Letter is required for access to PREDICT data, and that this letter must be signed by a person who has authority to act on behalf of the sponsoring institution. I have such authority. I understand that a Researcher may be sponsored individually by an organization or an organization may submit a Sponsorship Letter for the entity itself to function as a Researcher, naming one or more Data Custodians as the persons who will apply for PREDICT accounts and be responsible for the research and PREDICT data used by the entity. This letter is (check one):

☐ Sponsorship of Individual(s) as Researcher (if desired, more than one individual can be listed on a Sponsorship Letter if each Researcher would like to have a PREDICT account).

☐ Sponsorship of Entity as Researcher (if desired, more than one individual can be listed on a Sponsorship Letter as Data Custodian)

This letter is being sent on behalf of the following individual(s) or on behalf of the following Data Custodian(s) who will apply for PREDICT accounts on behalf of this organization:

Full Name (Name of Researcher or Name of Data Custodian)	**Organization Name**	**Department/ Business Unit**	**Title**	**Years With Org**

The above named persons have a legitimate need for PREDICT data, owing to their position within their department or business unit, and the responsibilities assigned to them.

If this is a Sponsorship of Individuals as Researcher(s), I hereby confirm that the above individual(s) is/are currently affiliated with this organization, is/are in good standing, and serve in the capacity noted above. I, or any successor in my role, will inform the PCC if any of the above named individual(s) leave our organization or otherwise have changed circumstances calling into question or eliminating their need for PREDICT data.

If this is a Sponsorship of Entity, I hereby confirm that the above named individual(s) will serve as Data Custodian(s) for our organization and are trusted by this organization to be responsible for PREDICT data and research efforts on its behalf. I, or any successor in my role, will inform the PCC if any of the individual(s) named above as a Data Custodian is/are no longer in that role or is/are no longer in charge of the data and associated research effort. At the same time, we will provide information on the Data Custodian(s) who will replace the previous one(s).

This organization appreciates the importance of cyber security and the value of R&D efforts in this area, and we are pleased to support the PREDICT project through this Sponsorship Letter. Please let me know if you need any further information.

Sincerely,

_____ _____

Signature Date

_____

Name Printed or Typed

_____

Title or Position

_____ _____

Email Telephone

COVER LETTER
MEMORANDUM OF AGREEMENT

Thank you for your interest in using PREDICT datasets. In order for your application for PREDICT datasets to be considered, you must complete and sign the attached Memorandum of Agreement (MOA) and submit it to the PREDICT Coordinating Center (PCC).

Instructions:

1. Print the MOA.
2. Fill in requested information and complete Attachment A, as noted.
3. Complete the Contact Information form below
4. Sign the MOA and fax it to the PCC, Attn: Renee Karlsen, at **866.835.0255 (toll free)**. An executed copy will be returned to you for your files.

Questions regarding this MOA or your request for PREDICT datasets may be directed to Ms. Renee Karlsen at (919) 541-7115 or via email: PREDICT-contact@rti.org.

Contact Information For Person Signing Document

Name _____

Title _____

Organization _____

Address _____

City _____ State _____ Zip _____

Email _____

Phone _____

Fax _____

DHS Authority to Collect This Information: The Homeland Security Act of 2002 [Public Law 107-296, §302(4)] authorizes the Science and Technology Directorate to conduct "basic and applied research, development, demonstration, testing, and evaluation activities that are relevant to any or all elements of the Department, through both intramural and extramural programs." In exercising its responsibility under the Homeland Security Act, S&T is authorized to collect information, as appropriate, to support R&D related to improving the security of the homeland. **Principal Purpose:** DHS collects name, organization and title (if any), email address, home and/or work address, and telephone numbers for the purpose of contacting individuals regarding the PREDICT project and/or their involvement with PREDICT. **Routine Uses and Sharing:** Some of your information will be disclosed to PREDICT team members, such as data hosts, data providers, PREDICT contractors, the Predict Coordinating Center, the advisory board, and review board members to help us deliver requested PREDICT services and operate the PREDICT Web site and deliver the services you have requested. Unless you consent otherwise, this information will not be used for any purpose other than those stated above. However, DHS may release this information for an individual on a case-by-case basis as described in the DHS/ALL-002 System of Records Notice (SORN), which can be found at: www.dhs.gov/privacy. **Disclosure:** Furnishing this information is entirely voluntary; however, failure to furnish at least the minimum information required to register (to include full name and email address,) will prevent you from obtaining authorization to access system.

PRA Burden Statement: An agency may not conduct or sponsor an information collection and a person is not required to respond to this information collection unless it displays a current valid OMB control number and an expiration date. The control number for this collection is 1640-0012 and this form will expire on 08/31/2010. The estimated average time to complete this form is 45 minutes per respondent. If you have any comments regarding the burden estimate you can write to Department of Homeland Security, Science and Technology Directorate, Washington, DC 20528.

MEMORANDUM OF AGREEMENT
PROVIDER AND RESEARCHER

This Memorandum of Agreement ("MOA" or "Agreement") is between _____, a _____ corporation or entity having offices at _____ *or* an individual, _____, with address of _____ ("Researcher/User" for either corporation/entity or individual) and Research Triangle Institute ("RTI"), a North Carolina corporation having offices at 3040 Cornwallis Road, Research Triangle Park, NC 27709, collectively referred to as "the Parties."

RTI serves under contract to the United States Department of Homeland Security ("DHS") as the operator of the PREDICT Coordinating Center ("PCC"). References throughout this document to "PCC" shall be deemed to refer to RTI. References to the MOA Identification number ("MOA ID") assigned at the top left of each page of this document shall refer to this Agreement.

Recitals

The PCC supports the Protected Repository for the Defense of Infrastructure against Cyber Threats (PREDICT) project sponsored by the United States Department of Homeland Security (DHS).

The following eight types of organizations participate in PREDICT:

Department of Homeland Security (DHS)	Data Providers	Researchers/Users	Application Review Board (ARB)
PCC	Data Hosts	Sponsoring Institutions	Publication Review Board (PRB)

This Agreement consists of: the General Terms and Conditions and Attachment A and any subsequent Amendment(s) to Researcher/User Agreement, if executed.

The provisions of Attachment A shall be construed so as to be fully consistent with all of the provisions of the General Terms and Conditions of this Agreement and, in the case of any conflict, the General Terms and Conditions shall prevail unless an Attachment is separately executed by both Parties and expressly amends particular provisions of the General Terms and Conditions, in which case the amendments of such Attachment shall prevail over such particular provisions of the General Terms and Conditions.

General Terms and Conditions

Researcher/User and PCC agree to the following:

Data shall mean the datasets described in Attachment A that are owned or controlled by a Data Provider, which are being requested by the Researcher/User.

Metadata is descriptive information about the Data (but not the Data itself) that is inserted in the PREDICT data catalog and serves as a description of the Data.

DHS shall mean the U.S. Department of Homeland Security.

PCC shall mean the Predict Coordinating Center that manages the PREDICT data catalog and operations, processes applications for PREDICT data, and handles requests for approval of publication and other administrative matters. PCC does not store, maintain, or have access to any of the Data.

Data Provider shall mean an entity that provides Data to the PREDICT project that it owns or has a right to control and disclose to the PREDICT project, subject to the terms and conditions in an MOA between it and PCC.

Data Host shall mean an entity that provides computing infrastructure to store Data received from one or more Data Providers, and provides approved Researchers/Users access to Data.

Researcher/User shall mean an approved person or entity that is requesting Data from PREDICT for use in research and who is responsible for the receipt, security, oversight, use, and return of Data.

Data Custodian shall mean the person designated by an entity Researcher/User who has primary responsibility for the receipt, security, oversight, use, and return of Data on behalf of a Researcher/User that is an entity, not an individual.

Sponsoring Institution is an organization that is affiliated with or otherwise sponsors Researchers/Users and validates their research and need for PREDICT data.

Application Review Board ("ARB") shall mean an entity that reviews and approves or rejects applications for (a) accounts to access the catalog of Metadata and (b) MOAs serving as applications for requested Data.

Publication Review Board ("PRB") shall mean an entity that reviews and comments upon applications from Researchers/Users to publish or otherwise release any study results or other information relating to research using Data received through PREDICT.

Researcher/User must sign Attachment A as a condition precedent to obtaining any Data under PREDICT.

Researcher/User Agreements, Rights, and Obligations

In consideration of the release to Researcher/User of the Data described in Attachment A, the Researcher/User agrees to the following terms and conditions:

1. Researcher/User certifies that all information provided by Researcher/User in this Agreement is accurate and complete.

2. Researcher/User agrees that all information contained in this Agreement may be shared as necessary to facilitate PCC operations and comply with PCC operational policies and procedures, including the sharing of information in this Agreement with the ARB, PRB, Data Hosts, Data Providers, and, if necessary, DHS.

3. Researcher/User agrees to use the Data solely for the research purpose described in Attachment A and in all respects in accordance with this Agreement (including the terms and terms and conditions specified in Attachment A).

4. Upon receipt of the Data, Data Provider hereby grants to Researcher/User, a license to Researcher/User to use the Data solely for the research purpose described in the Researcher/User's application.

5. Researcher/User agrees that he/she shall not transmit, send, export, or use the Data outside of the United States, and Researcher/User shall take steps to ensure that all persons named on Researcher/User's application are aware of this restriction and do not transmit, send, export, or use the Data outside the United States.

6. The Researcher/User shall not allow access to or use of Data to any persons other than those identified in Attachment A of this Agreement. Researcher/User shall initiate an Amendment to this Agreement if individuals other than those identified in Attachment A are to be given access to the Data. Such Amendment must be approved and signed by both the Researcher/User and the PCC prior to any new individuals being given access to any Data.

7. Researcher/User shall establish and maintain the appropriate administrative, technical, and physical safeguards to protect the confidentiality of the Data and to prevent unauthorized use or access to the Data. At a minimum, Researcher/User shall use at least the same degree of care in safeguarding Data he/she uses for his/her own proprietary information, provided such degree of care is reasonably calculated to prevent inadvertent disclosure or unauthorized use.

8. Researcher/User, if an individual, shall notify the PCC in writing within thirty (30) days if he/she leaves the Sponsoring Institution or the research project or, in the case of a Researcher/User that is an entity, if the Data Custodian is no longer serving in this capacity.

9. (a) Researcher/User as an Individual
If Researcher/User is an individual and (i) moves to a different institution after access to Data is granted, (ii) moves to another area of the Sponsoring Institution or for any other reason is no longer affiliated with the research associated with the Data, or (iii) dies, Researcher/User's approval to use or disclose the Data shall immediately be suspended, as shall use of Data by any other individual whether or not named in Researcher/User's application or located at Researcher/User's Sponsoring Institution. Researcher/User, or designate in the event of death, shall notify the PCC in writing within thirty (30) days of such event regarding the proposed disposition of all copies of the Data and follow PCC's directions as provided. Continued use of the Data to which Researcher/User had approved access shall be contingent upon the submission and approval of a new application for use of the datasets

(b) Researcher/User as an Entity
If Researcher/User is an entity and the individual identified as the Data Custodian (i) leaves employment with Researcher/User, (ii) moves to another area of the Sponsoring Institution or for any other reason is no longer affiliated with the research associated with the Data, or (iii) dies, Researcher/User (the entity) shall provide the PCC with an interim point of contact and propose a substitute Data Custodian to the PCC within thirty (30) days of such event via an Amendment to Researcher/User Agreement. PCC shall approve or deny the proposed substitution within five business days, or such longer period as may be required to obtain adequate information and/or approvals from Researcher/User or third parties as is necessary to fully evaluate the proposed Data Custodian's fitness for the position. During the period of review, no individuals other than those previously approved shall have access to the Data pending the PCC's decision.

10. No findings, analysis, or information derived from the Data may be released if such findings contain any combination of data elements that might allow for identification or the deduction of a person's or institution's identity, unless such identification is both (a) explicitly permitted under the terms governing handling and release of Data incorporated herein and (b) not in violation of applicable U.S. or state law.

11. Researcher/User shall submit any findings, results of analysis, or manuscripts proposed for public release, publication, or any other type of disclosure ("Writings") to persons not listed in Attachment A to the PCC for review and approval by a Publications Review Board (PRB). Writings involving confidential or proprietary information may be submitted to the PCC for review solely by the Data Provider(s) of the dataset(s) that was/were used and subject to specific non-disclosure provisions, including restrictions that the reviewer may not use any of the information under review for its own benefit or research without written permission from the Researcher/User. If the writings involve any U.S. Government classified information, such review shall be limited to review by the Data Provider if the Data Provider has appropriate clearances or, alternatively, by the DHS PREDICT Program Manager or other DHS personnel holding appropriate clearances. No further PRB review shall be required for Writings involving confidential or proprietary information or U.S. Government classified information. Researcher/User shall submit such Writings to the PCC at the same time that Researcher/User submits the Writings for conference or journal acceptance or for any other purpose. PRB review is limited to ensuring that data confidentiality is maintained, entities or individuals cannot be identified (except as permitted under Article 11 above), and the terms and conditions attached to the use of the Data have been followed.

12. Researcher/User shall identify the PREDICT program as the source of Data in all Writings and DHS as the sponsor of PREDICT. Researcher/User shall abide by any decisions made by the PCC and PRB with respect

to non-publication or changes necessary to ensure the conditions associated with the Data are met. Researcher/User shall not permit publication or otherwise publicly release such Writings until PRB approval has been received from the PCC. PCC may withhold approval to publish on the results of research only if it reasonably determines that the format of Data presentation is such that it does not meet the terms and conditions for the use of the Data as reflected in this MOA and Attachments or if publication of the Writings may result in identification of the Data Provider or another institution, organization, or individual or otherwise breach a duty of confidence owed to Data Provider or the subject from whom the Data was collected. Researcher/User shall provide to the PCC a link to or copy of all published Writings.

13. Researcher/User shall report immediately to PCC any use or disclosure of the Data other than as permitted by this Agreement. Researcher/User shall take all commercially reasonable steps to mitigate the effects of such improper use or disclosure, including cooperating with all reasonable requests of PCC.

14. Unless re-identification of Data is required and was disclosed and approved in the application by Researcher/User to the ARB, Researcher/User shall not attempt to or actually unlock, override, reverse engineer, or otherwise take any steps to defeat any anonymization or obfuscation methods or tools that have been applied to any Data by the Data Provider or Data Host, or otherwise to violate any of the terms of use associated with the Data.

15. Researcher/User agrees that in the event PCC determines or has a reasonable belief that Researcher/User has violated any terms of this Agreement, PCC may terminate this Agreement and require that Researcher/User destroy the Data and all derivative files pursuant to PCC instructions. PCC may also seek injunctive relief against Researcher/User to prevent any unauthorized disclosure of Data by Researcher/User. Researcher/User understands that as a result of this determination or reasonable belief that a violation of this Agreement has occurred, PCC may also refuse to release further Data to Researcher/User. In addition, PCC may report any misuse or improper disclosure of Data to Data Provider and Data Host and to appropriate authorities as permitted or required by applicable Federal or state law.

16. Access to Data ends upon expiration or termination of this Agreement and Researcher/User shall, as directed by PCC, destroy all copies of the Data per PCC's instructions. Researcher/User or the Data Custodian shall certify such destruction or return by signing and providing to PCC a Certification of Data Destruction.

17. Researcher/User shall be responsible for harm directly caused by the gross negligence or willful misconduct of Researcher/User or its agents, arising out of or connected to the use of any of the Data or research related to this Agreement.

18. Researcher/User shall promptly notify PCC of any claim against it or a third party of which it becomes aware pertaining to Data or research related to this Agreement.

PREDICT Coordinating Center (PCC) Rights and Obligations

1. PCC shall notify Researcher/User of
    a) Freedom of Information Act ("FOIA") or other legal requests for access to data regarding this Agreement; and
    b) Data destruction requirements at expiration of this Agreement.

2. PCC shall obtain from all Data Providers written agreement that (i) its Data complies with all restrictions specified by the PCC and all requirements of applicable governing or regulating bodies and/or contractual agreements, and (ii) that all Data is consistent with Data Provider's privacy, security, or other policies and procedures applicable to the Data.

3. PCC may terminate this Agreement upon determination that information provided in this Agreement was false, inaccurate, incomplete, or otherwise designed to conceal material information and require destruction of all Data (and copies thereof) that was provided to Researcher/User.

Joint Rights and Obligations – Researcher/User and PCC, and Other Provisions

1. Either party may terminate this Agreement by providing thirty (30) days written notice. Upon any termination or the expiration of this Agreement, Researcher/User shall, upon direction from PCC, destroy all copies of Data, or portions thereof, in its possession that it has received from Data Host or created (or had others create). Researcher/User shall certify (or in the case of an entity that is a Researcher/User, the Data Custodian shall certify) to PCC such destruction or return by signing and providing a Certification of Data Destruction.

2. This Agreement shall remain in force for a period of one year commencing with the date of latest signature below, or as amended. All obligations or rights, which by their nature survive and continue after the end date of this Agreement, shall survive and continue, and this shall specifically include the obligation of Researcher/User to seek review by the PCC and PRB prior to publication as noted above. Any Amendments to this Agreement, to be effective, shall be in writing and signed by an authorized Representative of each Party.

3. This Agreement shall be construed and interpreted in accordance with the laws of the state of North Carolina.

4. Nothing contained herein shall be construed as conferring by implication, estoppel or otherwise any license or right in favor of either party or any third party in any patents or other intellectual property rights of the other.

5. Neither Party shall in any manner reference or cause to be referenced the trade names, trademarks, service marks or any other indicia of origin owned by the other Party, or indicate that its operations are any way sponsored, approved or endorsed by the other.

6. Except with the written consent of Researcher/User, PCC shall not cause to be issued or released for publication, or participate in the publication of, any articles or publicity relating to Researcher/User and the subject matter of this Agreement; provided, however, that PCC may reveal all information supplied by Researcher/User in this MOA and its applications to the ARB or PRB, so long as that information is used only for purposes of evaluating those applications.

7. Neither this Agreement nor the receipt of Data by Researcher/User shall constitute or imply any promise or intention by Researcher/User to evaluate, process or make use of the Data either now or in the future.

8. **NEITHER PARTY SHALL BE LIABLE TO THE OTHER FOR INCIDENTAL, INDIRECT, CONSEQUENTIAL, SPECIAL, PUNITIVE, OR EXEMPLARY DAMAGES OF ANY KIND (INCLUDING LOST REVENUES OR PROFITS, OR LOSS OF BUSINESS) IN ANY WAY RELATED TO THIS AGREEMENT, REGARDLESS OF WHETHER IT WAS ADVISED, HAD OTHER REASON TO KNOW, OR IN FACT KNEW OF THE POSSIBILITY THEREOF.**

9. Any legal action arising in connection with this Agreement must begin within two (2) years after the cause of action arises.

10. The Parties may execute two or more copies of this Agreement, each of which shall constitute an original copy of this Agreement. A scanned, imaged, facsimile or photocopy of this Agreement or amendment to this Agreement as executed by the Parties shall be deemed to be an original executed copy for all purposes.

11. If Researcher/User is an individual, this Agreement shall not be considered accepted or effective until signed below by Researcher/User and the authorized representative of PCC. If Researcher/User is an entity, this Agreement shall not be considered accepted or effective until signed below by authorized representatives of both Parties, and each Party represents and warrants that the person signing this Agreement on its behalf has full authority to bind his or her organization to this Agreement. By signing below, neither Party may assign all or a portion of its rights and obligations hereunder without the prior written approval of the other Party.

RESEARCH TRIANGLE INSTITUTE **PREDICT Coordinating Center**		**RESEARCHER/USER**
Signature		Signature
Name		Name
Title		Title
Date		Date

Attachment A

Primary Researcher or Data Custodian if entity (Name, Organization, Address, Telephone, Email):

All Other Persons With Access to Datasets (Name, Organization, Address, Telephone, Email):

Proposed Use of Data:

Dataset(s) Requested:

[One row must be completed for each data set being requested in Researcher/User's Application]

Data Category	Dataset Name	IRB Approval Date (If Applicable)

Memorandum of Agreement
Attachment A - continued

Subject to the Following Additional Terms and Conditions for Access to and Use of Data as set by Data Provider

Subject to the Following Additional Terms and Conditions for Access to and Use of Data as set by Data Host (if any)

This Attachment A and its terms and conditions shall be a part of the Memorandum of Agreement between the PCC and Researcher/User upon both approval of Researcher/User's application for access to the requested datasets by the Application Review Board and signature below by Researcher/User. Such Agreement and signature shall be a condition precedent to Researcher/User's access to any Data requested in Attachment A.

RESEARCHER/USER
Signature
Name
Title
Date