

**02 INFORMATION ABOUT PRINCIPAL INVESTIGATORS/PROJECT DIRECTORS(PI/PD) and  
co-PRINCIPAL INVESTIGATORS/co-PROJECT DIRECTORS**

---

Submit only ONE copy of this form for each PI/PD and co-PI/PD identified on the proposal. The form(s) should be attached to the original proposal as specified in GPG Section II.C.a. Submission of this information is voluntary and is not a precondition of award. This information will not be disclosed to external peer reviewers. **DO NOT INCLUDE THIS FORM WITH ANY OF THE OTHER COPIES OF YOUR PROPOSAL AS THIS MAY COMPROMISE THE CONFIDENTIALITY OF THE INFORMATION.**

---

**PI/PD Name:** Hsinchun Chen

**Gender:**  Male  Female  
**Ethnicity:** (Choose one response)  Hispanic or Latino  Not Hispanic or Latino

**Race:**  
(Select one or more)  American Indian or Alaska Native  
 Asian  
 Black or African American  
 Native Hawaiian or Other Pacific Islander  
 White

**Disability Status:**  
(Select one or more)  Hearing Impairment  
 Visual Impairment  
 Mobility/Orthopedic Impairment  
 Other  
 None

**Citizenship:** (Choose one)  U.S. Citizen  Permanent Resident  Other non-U.S. Citizen

**Check here if you do not wish to provide any or all of the above information (excluding PI/PD name):**

**REQUIRED: Check here if you are currently serving (or have previously served) as a PI, co-PI or PD on any federally funded project**

---

**Ethnicity Definition:**

**Hispanic or Latino.** A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race.

**Race Definitions:**

**American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

**Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

**Black or African American.** A person having origins in any of the black racial groups of Africa.

**Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

**White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

---

**WHY THIS INFORMATION IS BEING REQUESTED:**

The Federal Government has a continuing commitment to monitor the operation of its review and award processes to identify and address any inequities based on gender, race, ethnicity, or disability of its proposed PIs/PDs. To gather information needed for this important task, the proposer should submit a single copy of this form for each identified PI/PD with each proposal. Submission of the requested information is voluntary and will not affect the organization's eligibility for an award. However, information not submitted will seriously undermine the statistical validity, and therefore the usefulness, of information received from others. Any individual not wishing to submit some or all the information should check the box provided for this purpose. (The exceptions are the PI/PD name and the information about prior Federal support, the last question above.)

Collection of this information is authorized by the NSF Act of 1950, as amended, 42 U.S.C. 1861, et seq. Demographic data allows NSF to gauge whether our programs and other opportunities in science and technology are fairly reaching and benefiting everyone regardless of demographic category; to ensure that those in under-represented groups have the same knowledge of and access to programs and other research and educational opportunities; and to assess involvement of international investigators in work supported by NSF. The information may be disclosed to government contractors, experts, volunteers and researchers to complete assigned work; and to other government agencies in order to coordinate and assess programs. The information may be added to the Reviewer file and used to select potential candidates to serve as peer reviewers or advisory committee members. See Systems of Records, NSF-50, "Principal Investigator/Proposal File and Associated Records", 63 Federal Register 267 (January 5, 1998), and NSF-51, "Reviewer/Proposal File and Associated Records", 63 Federal Register 268 (January 5, 1998).

**02 INFORMATION ABOUT PRINCIPAL INVESTIGATORS/PROJECT DIRECTORS(PI/PD) and  
co-PRINCIPAL INVESTIGATORS/co-PROJECT DIRECTORS**

---

Submit only ONE copy of this form for each PI/PD and co-PI/PD identified on the proposal. The form(s) should be attached to the original proposal as specified in GPG Section II.C.a. Submission of this information is voluntary and is not a precondition of award. This information will not be disclosed to external peer reviewers. **DO NOT INCLUDE THIS FORM WITH ANY OF THE OTHER COPIES OF YOUR PROPOSAL AS THIS MAY COMPROMISE THE CONFIDENTIALITY OF THE INFORMATION.**

---

**PI/PD Name:** Catherine A Larson

**Gender:**  Male  Female  
**Ethnicity:** (Choose one response)  Hispanic or Latino  Not Hispanic or Latino

**Race:**  
(Select one or more)  
 American Indian or Alaska Native  
 Asian  
 Black or African American  
 Native Hawaiian or Other Pacific Islander  
 White

**Disability Status:**  
(Select one or more)  
 Hearing Impairment  
 Visual Impairment  
 Mobility/Orthopedic Impairment  
 Other  
 None

**Citizenship:** (Choose one)  U.S. Citizen  Permanent Resident  Other non-U.S. Citizen

**Check here if you do not wish to provide any or all of the above information (excluding PI/PD name):**

**REQUIRED: Check here if you are currently serving (or have previously served) as a PI, co-PI or PD on any federally funded project**

---

**Ethnicity Definition:**

**Hispanic or Latino.** A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race.

**Race Definitions:**

**American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

**Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

**Black or African American.** A person having origins in any of the black racial groups of Africa.

**Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

**White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

---

**WHY THIS INFORMATION IS BEING REQUESTED:**

The Federal Government has a continuing commitment to monitor the operation of its review and award processes to identify and address any inequities based on gender, race, ethnicity, or disability of its proposed PIs/PDs. To gather information needed for this important task, the proposer should submit a single copy of this form for each identified PI/PD with each proposal. Submission of the requested information is voluntary and will not affect the organization's eligibility for an award. However, information not submitted will seriously undermine the statistical validity, and therefore the usefulness, of information received from others. Any individual not wishing to submit some or all the information should check the box provided for this purpose. (The exceptions are the PI/PD name and the information about prior Federal support, the last question above.)

Collection of this information is authorized by the NSF Act of 1950, as amended, 42 U.S.C. 1861, et seq. Demographic data allows NSF to gauge whether our programs and other opportunities in science and technology are fairly reaching and benefiting everyone regardless of demographic category; to ensure that those in under-represented groups have the same knowledge of and access to programs and other research and educational opportunities; and to assess involvement of international investigators in work supported by NSF. The information may be disclosed to government contractors, experts, volunteers and researchers to complete assigned work; and to other government agencies in order to coordinate and assess programs. The information may be added to the Reviewer file and used to select potential candidates to serve as peer reviewers or advisory committee members. See Systems of Records, NSF-50, "Principal Investigator/Proposal File and Associated Records", 63 Federal Register 267 (January 5, 1998), and NSF-51, "Reviewer/Proposal File and Associated Records", 63 Federal Register 268 (January 5, 1998).

**02 INFORMATION ABOUT PRINCIPAL INVESTIGATORS/PROJECT DIRECTORS(PI/PD) and  
co-PRINCIPAL INVESTIGATORS/co-PROJECT DIRECTORS**

---

Submit only ONE copy of this form for each PI/PD and co-PI/PD identified on the proposal. The form(s) should be attached to the original proposal as specified in GPG Section II.C.a. Submission of this information is voluntary and is not a precondition of award. This information will not be disclosed to external peer reviewers. **DO NOT INCLUDE THIS FORM WITH ANY OF THE OTHER COPIES OF YOUR PROPOSAL AS THIS MAY COMPROMISE THE CONFIDENTIALITY OF THE INFORMATION.**

---

**PI/PD Name:** Mark W Patton

**Gender:**  Male  Female  
**Ethnicity:** (Choose one response)  Hispanic or Latino  Not Hispanic or Latino

**Race:**  
(Select one or more)  
 American Indian or Alaska Native  
 Asian  
 Black or African American  
 Native Hawaiian or Other Pacific Islander  
 White

**Disability Status:**  
(Select one or more)  
 Hearing Impairment  
 Visual Impairment  
 Mobility/Orthopedic Impairment  
 Other  
 None

**Citizenship:** (Choose one)  U.S. Citizen  Permanent Resident  Other non-U.S. Citizen

**Check here if you do not wish to provide any or all of the above information (excluding PI/PD name):**

**REQUIRED: Check here if you are currently serving (or have previously served) as a PI, co-PI or PD on any federally funded project**

---

**Ethnicity Definition:**

**Hispanic or Latino.** A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race.

**Race Definitions:**

**American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

**Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

**Black or African American.** A person having origins in any of the black racial groups of Africa.

**Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

**White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

---

**WHY THIS INFORMATION IS BEING REQUESTED:**

The Federal Government has a continuing commitment to monitor the operation of its review and award processes to identify and address any inequities based on gender, race, ethnicity, or disability of its proposed PIs/PDs. To gather information needed for this important task, the proposer should submit a single copy of this form for each identified PI/PD with each proposal. Submission of the requested information is voluntary and will not affect the organization's eligibility for an award. However, information not submitted will seriously undermine the statistical validity, and therefore the usefulness, of information received from others. Any individual not wishing to submit some or all the information should check the box provided for this purpose. (The exceptions are the PI/PD name and the information about prior Federal support, the last question above.)

Collection of this information is authorized by the NSF Act of 1950, as amended, 42 U.S.C. 1861, et seq. Demographic data allows NSF to gauge whether our programs and other opportunities in science and technology are fairly reaching and benefiting everyone regardless of demographic category; to ensure that those in under-represented groups have the same knowledge of and access to programs and other research and educational opportunities; and to assess involvement of international investigators in work supported by NSF. The information may be disclosed to government contractors, experts, volunteers and researchers to complete assigned work; and to other government agencies in order to coordinate and assess programs. The information may be added to the Reviewer file and used to select potential candidates to serve as peer reviewers or advisory committee members. See Systems of Records, NSF-50, "Principal Investigator/Proposal File and Associated Records", 63 Federal Register 267 (January 5, 1998), and NSF-51, "Reviewer/Proposal File and Associated Records", 63 Federal Register 268 (January 5, 1998).

## List of Suggested Reviewers or Reviewers Not To Include (optional)

---

### **SUGGESTED REVIEWERS:**

- Antonio Badia; University of Louisville; [abadia@louisville.edu](mailto:abadia@louisville.edu)
- Sharad Mehrotra; University of California at Irvine; [sharad@ics.uci.edu](mailto:sharad@ics.uci.edu)
- H. Raghav Rao; UB School of Management; [mgmtrao@buffalo.edu](mailto:mgmtrao@buffalo.edu)
- Ramesh Sharda; Oklahoma State University; [ramesh.sharda@okstate.edu](mailto:ramesh.sharda@okstate.edu)
- Lina Zhou; University of Maryland, Baltimore County; [zhoul@umbc.edu](mailto:zhoul@umbc.edu)

### **REVIEWERS NOT TO INCLUDE:**

Not Listed

---

## CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research and Community

(Pilot Demonstration Project)

PI: H. Chen.

### Conflict of Interest List - Integrated and Ordered by COI Last Name

<u>PI/SP Last Name</u>	<u>PI/SP First Name</u>	<u>PI/SP Institution</u>	<u>COI Last Name</u>	<u>COI First Name</u>	<u>COI Institution</u>
Chen	Hsinchun	University of Arizona	Abbasi	Ahmed	University of Virginia
Abbasi	Ahmed	University of Virginia	Acharya	Subrata	Towson University
Abbasi	Ahmed	University of Virginia	Adjeroh	Donald	West Virginia University
Abbasi	Ahmed	University of Virginia	Albrecht	Conan	Brigham Young University
Thuraisingham	Bhavani	University of Texas at Dallas	Alipanah	Neda	UTD – former advisee
Yang	Chris	Drexel University	Ananiadou	Sophia	University of Manchester, UK)
Chen	Hsinchun	University of Arizona	Ascher	Michael	LLNL
Chen	Hsinchun	University of Arizona	Atabakhsh	Homa	Raytheon
Abbasi	Ahmed	University of Virginia	Benjamin	Victor	University of Arizona
Thuraisingham	Bhavani	University of Texas at Dallas	Bertino	E.	Purdue
Thuraisingham	Bhavani	University of Texas at Dallas	Bhargava	B.	Purdue
Hu	Paul	University of Utah	Bonner	Bryan	University of Utah
Chen	Hsinchun	University of Arizona	Brandt	Lawrence	National Science Foundation
Chen	Hsinchun	University of Arizona	Brown	Randall	UA Health Network
Hu	Paul	University of Utah	Brown	Sue	University of Arizona
Chen	Hsinchun	University of Arizona	Brown	Susan	University of Arizona
Thuraisingham	Bhavani	University of Texas at Dallas	Cadenhead	Tyrone	UTD – former advisee
Chen	Hsinchun	University of Arizona	Campbell	John	University of Georgia
Thuraisingham	Bhavani	University of Texas at Dallas	Carminati	B.	U of Como, U. of Milan - advisee
Yang	Chris	Drexel University	Chan	Mandy	Drexel Univ. – past advisee
Yang	Chris	Drexel University	Chan	Patricia	Drexel Univ. – past advisee
Yang	Chris	Drexel University	Chan	Winter	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Chau	Michael	Hong Kong University
Hu	Paul	University of Utah	Chau	Patrick	University of Hong Kong
Chen	Hsinchun	University of Arizona	Chen	Ching-chih	Simmons College
Hu	Paul	University of Utah	Chen	Hsinchun	University of Arizona
Abbasi	Ahmed	University of Virginia	Chen	Hsinchun	University of Arizona
Chen	Hsinchun	University of Arizona	Chen	Kuo-Tay	National Taiwan University

<u>PI/SP Last Name</u>	<u>PI/SP First Name</u>	<u>PI/SP Institution</u>	<u>COI Last Name</u>	<u>COI First Name</u>	<u>COI Institution</u>
Thuraisingham	Bhavani	University of Texas at Dallas	Chen	S.	Univ. of Minn. - advisee
Chen	Hsinchun	University of Arizona	Chen	Su-Shing	The University of Florida
Chen	Hsinchun	University of Arizona	Chen	Tsai-Jyh	National Chengchi University
Abbasi	Ahmed	University of Virginia	Chen	Yan	University of Auburn-Montgomery
Abbasi	Ahmed	University of Virginia	Chen	Yi Da	University of Arizona
Yang	Chris	Drexel University	Cheng	Tung-Yin	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Chiang	Roger	University of Cincinnati
Yang	Chris	Drexel University	Chuang	Katherine	Drexel Univ. – past advisee
Yang	Chris	Drexel University	Chung	Alan	Drexel Univ. – past advisee
Abbasi	Ahmed	University of Virginia	Chung	Wingyan	UNC-Fayetteville
Chen	Hsinchun	University of Arizona	Chung	Wingyan	Santa Clara University
Thuraisingham	Bhavani	University of Texas at Dallas	Clifton	C.	Purdue
Yang	Chris	Drexel University	Combi	Carlo	Universita' degli Studi di Verona, Italy
Larson	Catherine	University of Arizona	Dang	Mandy	N. Arizona University
Chen	Hsinchun	University of Arizona	Dang	Yan (Mandy)	Northern Arizona University
Thuraisingham	Bhavani	University of Texas at Dallas	Das	S.	NSF/UTA
Chen	Hsinchun	University of Arizona	Dawes	Sharon	University of Albany
Chen	Hsinchun	University of Arizona	Demchak	Chris	US Naval War College
Chen	Hsinchun	University of Arizona	Denning	Dorothy	Naval Postgraduate School
Chen	Hsinchun	University of Arizona	Dhar	Vasant	New York University
Chen	Hsinchun	University of Arizona	Dugan	Laura	Univ. of Maryland
Abbasi	Ahmed	University of Virginia	Duvall	Bob	West Virginia University
Chen	Hsinchun	University of Arizona	Eidson	Millicent	NY Dept. of Health
Larson	Catherine	University of Arizona	Eidson	Millicent	NY Dept. of Health
Yang	Chris	Drexel University	Faculty from Drexel		Drexel Univ. Faculty
Hu	Paul	University of Utah	Faculty from U. Utah		Faculty from University of Utah
Thuraisingham	Bhavani	University of Texas at Dallas	Faculty from UTD		UT-Dallas Faculty
Hu	Paul	University of Utah	Fang	Xiao	University of Utah
Thuraisingham	Bhavani	University of Texas at Dallas	Ferrari	E.	U of Como
Thuraisingham	Bhavani	University of Texas at Dallas	Finin	T.	UMBC
Chen	Hsinchun	University of Arizona	France	Stephen	University of Wisconsin, Milwaukee
Abbasi	Ahmed	University of Virginia	France	Stephen	Univ. of Wisconsin-Milwaukee
Chen	Hsinchun	University of Arizona	Friedman	Carol	Columbia University

<u>PI/SP Last Name</u>	<u>PI/SP First Name</u>	<u>PI/SP Institution</u>	<u>COI Last Name</u>	<u>COI First Name</u>	<u>COI Institution</u>
Abbasi	Ahmed	University of Virginia	Fu	Tianjun	Google Inc.
Chen	Hsinchun	University of Arizona	Fuller	Sherrilynne	University of Washington
Chen	Hsinchun	University of Arizona	Ganor	Boaz	Interdisciplinary Center Herzliya
Chen	Hsinchun	University of Arizona	Ganor	Boaz	Interdisciplinary Center Herzliya
Yang	Chris	Drexel University	Glaesser	Uwe	Simon Fraser U., Canada
Chen	Hsinchun	University of Arizona	Gotham	Ivan	NY Dept. of Health
Larson	Catherine	University of Arizona	Gotham	Ivan	NY Dept. of Health
Thuraisingham	Bhavani	University of Texas at Dallas	Han	J.	UIUC
Abbasi	Ahmed	University of Virginia	Hansen	James V.	Brigham Young University
Abbasi	Ahmed	University of Virginia	Hassan	Ammar	University of Virginia
Chen	Hsinchun	University of Arizona	Hersh	William	Oregon Health & Science Univ.
Yang	Chris	Drexel University	Hong	Kay	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Houston	Andrea	Louisiana State University
Chen	Hsinchun	University of Arizona	Hovy	Edward	University of Southern California
Hu	Paul	University of Utah	Hu	Han-fen	University of Nevada-Las Vegas
Hu	Paul	University of Utah	Hu	Han-fen	Advisee--University of Utah (2012) / UNLV
Chen	Hsinchun	University of Arizona	Hu	Paul	University of Utah
Larson	Catherine	University of Arizona	Hu	Paul	U. of Utah
Chen	Hsinchun	University of Arizona	Huang	Chun-Neng	Microsoft
Chen	Hsinchun	University of Arizona	Huang	Zan	Penn State University
Chen	Hsinchun	University of Arizona	Huang	Zan	Penn State University
Chen	Hsinchun	University of Arizona	Hubbard	Susan	NIH National Cancer Institute
Chen	Hsinchun	University of Arizona	Hung	Mao-Wei	National Taiwan University
Thuraisingham	Bhavani	University of Texas at Dallas	Hwang	K.	Univ. of Minn. - advisee
Thuraisingham	Bhavani	University of Texas at Dallas	Iyengar	S.	LSU
Thuraisingham	Bhavani	University of Texas at Dallas	Jajodia	S.	GMU
Thuraisingham	Bhavani	University of Texas at Dallas	Joshi	A.	UMBC
Chen	Hsinchun	University of Arizona	Kaza	Siddharth	Towson University
Abbasi	Ahmed	University of Virginia	Kaza	Siddharth	Towson University
Thuraisingham	Bhavani	University of Texas at Dallas	Keefe	T.	Univ. of Minn. - advisee
Thuraisingham	Bhavani	University of Texas at Dallas	Khadilkar	Vaibhav	UTD - advisee
Chen	Hsinchun	University of Arizona	Kiang	Melody	California State University
Hu	Paul	University of Utah	Kim	Iljoo	Advisee--University of Utah

<u>PI/SP Last Name</u>	<u>PI/SP First Name</u>	<u>PI/SP Institution</u>	<u>COI Last Name</u>	<u>COI First Name</u>	<u>COI Institution</u>
Thuraisingham	Bhavani	University of Texas at Dallas	Kim	K.	UC Irvine
Larson	Catherine	University of Arizona	Kruger	Betsy	UIUC
Chen	Hsinchun	University of Arizona	Lally	Ann	University of Washington
Yang	Chris	Drexel University	Lam	Wai	Chinese U. of Hong Kong, HK
Abbasi	Ahmed	University of Virginia	Larson	Cathy	University of Arizona
Thuraisingham	Bhavani	University of Texas at Dallas	Layfield	R.	UTD – former advisee
Chen	Hsinchun	University of Arizona	Li	Jiexun Jason	Drexel University
Yang	Chris	Drexel University	Li	Kar-Wing	Drexel Univ. – past advisee
Hu	Paul	University of Utah	Li	Lionel	Advisee--University of Utah
Chen	Hsinchun	University of Arizona	Li	Shu-Shing	National Taiwan University
Chen	Hsinchun	University of Arizona	Li	Xin	Hong Kong University
Larson	Catherine	University of Arizona	Li	Xin	Hong Kong University
Chen	Hsinchun	University of Arizona	Li	Yijun	Harbin Institute of Technology
Chen	Hsinchun	University of Arizona	Lim	Ee-Peng	Singapore Management Univ.
Yang	Chris	Drexel University	Lim	Ee-Peng	Singapore Management U., Singapore
Yang	Chris	Drexel University	Lin	Jianfeng	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Ling	Chienting	Pace University
Yang	Chris	Drexel University	Liu	Jiming	Hong Kong Baptist University
Thuraisingham	Bhavani	University of Texas at Dallas	Liu	Li	UTD – former advisee
Yang	Chris	Drexel University	Liu	Nan	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Lu	Hsinmin	National Taiwan University
Yang	Chris	Drexel University	Lu	Zhiyong	National Library of Medicine
Yang	Chris	Drexel University	Luk	Johnny	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Lynch	Cecil	Calif. Dept. of Health
Chen	Hsinchun	University of Arizona	Macintosh	Ann	Leeds University
Chen	Hsinchun	University of Arizona	Marshall	Byron	Oregon State University
Chen	Hsinchun	University of Arizona	McDonald	Dan	University of Utah
Thuraisingham	Bhavani	University of Texas at Dallas	Members of Mitre Corp.		MITRE Corp. members
Yang	Chris	Drexel University	Naikan	V.	Drexel Univ. – current advisee
Yang	Chris	Drexel University	Ng	Anthony	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Ng	Dorbin	Hong Kong University
Yang	Chris	Drexel University	Ng	Tobun D.	Drexel Univ. – current advisee
Chen	Hsinchun	University of Arizona	Nunamaker	Jay	University of Arizona



<u>PI/SP Last Name</u>	<u>PI/SP First Name</u>	<u>PI/SP Institution</u>	<u>COI Last Name</u>	<u>COI First Name</u>	<u>COI Institution</u>
Abbasi	Ahmed	University of Virginia	Nunamaker Jr.	Jay F.	University of Arizona
Thuraisingham	Bhavani	University of Texas at Dallas	Pallabi	Parveen	UTD – former advisee
Thuraisingham	Bhavani	University of Texas at Dallas	Partyka	Jeff	UTD – former advisee
Chen	Hsinchun	University of Arizona	Petersen	Timothy	Tucson Police Dept.
Abbasi	Ahmed	University of Virginia	Pinto	Santiago	West Virginia University
Chen	Hsinchun	University of Arizona	Qin	Jialun	Univ. of Massachusetts-Lowell
Chen	Hsinchun	University of Arizona	Quiñones	Karin	Southern Methodist University
Thuraisingham	Bhavani	University of Texas at Dallas	Rachapalli	Jyothsna	UTD - advisee
Chen	Hsinchun	University of Arizona	Raghu	Santanam	Arizona State University
Chen	Hsinchun	University of Arizona	Ramesh	Ramaswamy	SUNY Buffalo
Chen	Hsinchun	University of Arizona	Ramsey	Marshall	Microsoft
Chen	Hsinchun	University of Arizona	Reid	Edna	Federal Bureau of Investigation
Chen	Hsinchun	University of Arizona	Roberts	Nancy	Naval Postgraduate School
Chen	Hsinchun	University of Arizona	Roco	Mihail	National Science Foundation
Larson	Catherine	University of Arizona	Roco	Mihail	NSF
Chen	Hsinchun	University of Arizona	Romano	Nicolas	University of Tulsa
Abbasi	Ahmed	University of Virginia	Ross	Arun	West Virginia University
Chen	Hsinchun	University of Arizona	Roussinov	Dimitri	Arizona State University
Chen	Hsinchun	University of Arizona	Ryan	Johnny	IIEA (Ireland)
Chen	Hsinchun	University of Arizona	Sageman	Marc	Private consultant
Chen	Hsinchun	University of Arizona	Sageman	Marc	Private consultant
Thuraisingham	Bhavani	University of Texas at Dallas	Sahni	S.	UF
Chen	Hsinchun	University of Arizona	Salem	Arab	Consultant
Thuraisingham	Bhavani	University of Texas at Dallas	Sandhu	R.	GMU
Chen	Hsinchun	University of Arizona	Scanlon	Pamela	ARJIS (San Diego, CA)
Chen	Hsinchun	University of Arizona	Scanlon	Pamela	ARJIS (San Diego, CA)
Chen	Hsinchun	University of Arizona	Schatz	Bruce	University of Illinois
Chen	Hsinchun	University of Arizona	Schroeder	Jenny	Tucson Police Dept.
Chen	Hsinchun	University of Arizona	Schumaker	Rob	Cleveland State University
Chen	Hsinchun	University of Arizona	Sewell	Robin	University of Arizona
Thuraisingham	Bhavani	University of Texas at Dallas	She	Wei	UTD – former advisee
Thuraisingham	Bhavani	University of Texas at Dallas	Shekhar	S.	UMN
Hu	Paul	University of Utah	Sheng	Olivia	University of Utah - advisor

<u>PI/SP Last Name</u>	<u>PI/SP First Name</u>	<u>PI/SP Institution</u>	<u>COI Last Name</u>	<u>COI First Name</u>	<u>COI Institution</u>
Yang	Chris	Drexel University	Shi	Xiaodong	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Silke	Andrew	University of East London, U.K.
Chen	Hsinchun	University of Arizona	Sinai	Joshua	The Analysis Corporation
Chen	Hsinchun	University of Arizona	Son	Jaebong	University of Arizona
Chen	Hsinchun	University of Arizona	Steele	Logan	University of Connecticut
Chen	Hsinchun	University of Arizona	Storey	Veda	Georgia State University
Chen	Hsinchun	University of Arizona	Suakkaphong	Nichalin	University of Massachusetts, Dartmouth
Yang	Chris	Drexel University	Tang	Xuning	Drexel Univ. – past advisee
Yang	Chris	Drexel University	Theng	Yin-Leng	Nanyang Technological University, Singapore
Yang	Chris	Drexel University	Thuraisingham	Bhavani	U. of Texas, Dallas, USA
Chen	Hsinchun	University of Arizona	Thurmond	Mark	UC Davis
Chen	Hsinchun	University of Arizona	Tolle	Kris	Microsoft
Chen	Hsinchun	University of Arizona	Traunmuller	Roland	University of Linz (Austria)
Chen	Hsinchun	University of Arizona	Tsai	Feng-Tse	Asia University
Chen	Hsinchun	University of Arizona	UA Faculty		University of Arizona Faculty
Larson	Catherine	University of Arizona	UA Faculty		University of Arizona Faculty
Patton	Mark	University of Arizona	UA Faculty		University of Arizona Faculty
Yang	Chris	Drexel University	Unger	Jennifer	University of Southern California
Abbasi	Ahmed	University of Virginia	Vance	Anthony	Brigham Young University
Chen	Hsinchun	University of Arizona	Vinze	Ajay	Arizona State University
Chen	Hsinchun	University of Arizona	Violette	Charles	Tucson Police Dept.
Yang	Chris	Drexel University	Wang	Fu-Lee	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Wang	Gang Alan	Virginia Tech
Chen	Hsinchun	University of Arizona	Wang	Jau-Hwang	Central Police University
Yang	Chris	Drexel University	Wang	Ke	Simon Fraser U., Canada
Thuraisingham	Bhavani	University of Texas at Dallas	Wang	Zhong	UTD – former advisee
Hu	Paul	University of Utah	Wei	Chih-Ping	National Taiwan University
Yang	Chris	Drexel University	Wei	Chih-Ping	National Tsing Hua U., Taiwan
Chen	Hsinchun	University of Arizona	Weimann	Gabriel	Haifa University (Israel)
Chen	Hsinchun	University of Arizona	Wilson	Pete	Pima County Sheriff's Dept.
Thuraisingham	Bhavani	University of Texas at Dallas	Winslett	M.	UIUC
Yang	Chris	Drexel University	Winston	Flaura	U. of Pennsylvania, USA
Yang	Chris	Drexel University	Wong	Rachel	Drexel Univ. – past advisee

<u>PI/SP Last Name</u>	<u>PI/SP First Name</u>	<u>PI/SP Institution</u>	<u>COI Last Name</u>	<u>COI First Name</u>	<u>COI Institution</u>
Yang	Chris	Drexel University	Wong	Ying-Chau	Drexel Univ. – past advisee
Chen	Hsinchun	University of Arizona	Woo	Ji-Young	Korea
Chen	Hsinchun	University of Arizona	Wyzga	Wojciech	Knowledge Computing Corp.
Chen	Hsinchun	University of Arizona	Xu	Jie Jennifer	Bentley College
Chen	Hsinchun	University of Arizona	Yang	Christopher	Drexel University
Chen	Hsinchun	University of Arizona	Yang	Ming	Harbin Institute of Technology
Thuraisingham	Bhavani	University of Texas at Dallas	Yau	S.	ASU
Yang	Chris	Drexel University	Yen	John	Pennsylvania State University
Yang	Chris	Drexel University	Yip	Milo	Drexel Univ. – past advisee
Hu	Paul	University of Utah	Yu	Jongtae	Advisee--University of Utah
Yang	Chris	Drexel University	Yung	Stanley	Drexel Univ. – past advisee
Abbasi	Ahmed	University of Virginia	Zahedi	Mariam	University of Wisconsin-Milw.
Yang	Chris	Drexel University	Zeng	Daniel	U. of Arizona, USA
Abbasi	Ahmed	University of Virginia	Zeng	Daniel	University of Arizona
Yang	Chris	Drexel University	Zeng	Qing	University of Utah)
Chen	Hsinchun	University of Arizona	Zeng	Shuo	University of Arizona
Chen	Hsinchun	University of Arizona	Zhang	Yulei	Northern Arizona University
Larson	Catherine	University of Arizona	Zhang	Yulei	N. Arizona University
Abbasi	Ahmed	University of Virginia	Zhang	Zhu	University of Arizona
Chen	Hsinchun	University of Arizona	Zhou	Yilu	George Washington University
Chen	Hsinchun	University of Arizona	Zhu	Bin	Boston University
Yang	Chris	Drexel University	Zhu	Shanfeng	Drexel Univ. – current advisee
Abbasi	Ahmed	University of Virginia	Zimbra	David	Santa Clara University
Chen	Hsinchun	University of Arizona	Zimbra	David	Santa Clara University

## COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE/if not in response to a program announcement/solicitation enter NSF 14-1					<b>FOR NSF USE ONLY</b>	
NSF 14-530			04/09/14		<b>NSF PROPOSAL NUMBER</b>	
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.)					<b>1443019</b>	
ACI - DATANET						
DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION	
04/09/2014	1	05090000 ACI	7726	806345617	04/09/2015 4:07pm S	
EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)		SHOW PREVIOUS AWARD NO. IF THIS IS <input type="checkbox"/> A RENEWAL <input type="checkbox"/> AN ACCOMPLISHMENT-BASED RENEWAL		IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES <input type="checkbox"/> NO <input checked="" type="checkbox"/> IF YES, LIST ACRONYM(S)		
742652689						
NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE			ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE			
University of Arizona			888 N Euclid Ave TUCSON, AZ 85721-0001			
AWARDEE ORGANIZATION CODE (IF KNOWN)						
0010835000						
NAME OF PRIMARY PLACE OF PERF			ADDRESS OF PRIMARY PLACE OF PERF, INCLUDING 9 DIGIT ZIP CODE			
Univ. of Arizona MIS Dept.			Univ. of Arizona MIS Dept. 1130 E. Helen St., Rm 430 Tucson ,AZ ,857210108 ,US.			
IS AWARDEE ORGANIZATION (Check All That Apply) (See GPG II.C For Definitions)		<input type="checkbox"/> SMALL BUSINESS <input type="checkbox"/> FOR-PROFIT ORGANIZATION		<input type="checkbox"/> MINORITY BUSINESS <input type="checkbox"/> WOMAN-OWNED BUSINESS		<input type="checkbox"/> IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE
TITLE OF PROPOSED PROJECT <b>CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research and Community</b>						
REQUESTED AMOUNT \$	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE			
1,499,531	36 months	08/01/14				
THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW						
<input type="checkbox"/> BEGINNING INVESTIGATOR (GPG I.G.2)			<input type="checkbox"/> HUMAN SUBJECTS (GPG II.D.7) Human Subjects Assurance Number _____ Exemption Subsection _____ or IRB App. Date _____			
<input type="checkbox"/> DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C.1.e)			<input type="checkbox"/> INTERNATIONAL ACTIVITIES: COUNTRY/COUNTRIES INVOLVED (GPG II.C.2.j)			
<input type="checkbox"/> PROPRIETARY & PRIVILEGED INFORMATION (GPG I.D, II.C.1.d)						
<input type="checkbox"/> HISTORIC PLACES (GPG II.C.2.j)						
<input type="checkbox"/> VERTEBRATE ANIMALS (GPG II.D.6) IACUC App. Date _____ PHS Animal Welfare Assurance Number _____			<input checked="" type="checkbox"/> COLLABORATIVE STATUS			
<input checked="" type="checkbox"/> FUNDING MECHANISM <b>Research - other than RAPID or EAGER</b>			<b>Not a collaborative proposal</b>			
PI/PD DEPARTMENT		PI/PD POSTAL ADDRESS				
Management Information Systems		1130 E Helen St McClelland Hall, Rm 430Z Tucson, AZ 85721 United States				
PI/PD FAX NUMBER						
520-621-2433						
NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Email Address		
PI/PD NAME						
Hsinchun Chen	PhD	1989	520-621-4153	hchen@eller.arizona.edu		
CO-PI/PD						
Catherine A Larson	MS	1985	602-621-2211	cal@eller.arizona.edu		
CO-PI/PD						
Mark W Patton	MBA	1999	520-626-8614	mpatton@eller.arizona.edu		
CO-PI/PD						
CO-PI/PD						

## CERTIFICATION PAGE

### Certification for Authorized Organizational Representative (or Equivalent) or Individual Applicant

By electronically signing and submitting this proposal, the Authorized Organizational Representative (AOR) or Individual Applicant is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding conflict of interest (when applicable), drug-free workplace, debarment and suspension, lobbying activities (see below), nondiscrimination, flood hazard insurance (when applicable), responsible conduct of research, organizational support, Federal tax obligations, unpaid Federal tax liability, and criminal convictions as set forth in the NSF Proposal & Award Policies & Procedures Guide, Part I: the Grant Proposal Guide (GPG). Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U.S. Code, Title 18, Section 1001).

### Certification Regarding Conflict of Interest

The AOR is required to complete certifications stating that the organization has implemented and is enforcing a written policy on conflicts of interest (COI), consistent with the provisions of AAG Chapter IV.A.; that, to the best of his/her knowledge, all financial disclosures required by the conflict of interest policy were made; and that conflicts of interest, if any, were, or prior to the organization's expenditure of any funds under the award, will be, satisfactorily managed, reduced or eliminated in accordance with the organization's conflict of interest policy. Conflicts that cannot be satisfactorily managed, reduced or eliminated and research that proceeds without the imposition of conditions or restrictions when a conflict of interest exists, must be disclosed to NSF via use of the Notifications and Requests Module in FastLane.

### Drug Free Work Place Certification

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent), is providing the Drug Free Work Place Certification contained in Exhibit II-3 of the Grant Proposal Guide.

### Debarment and Suspension Certification

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

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency?

Yes

No

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) or Individual Applicant is providing the Debarment and Suspension Certification contained in Exhibit II-4 of the Grant Proposal Guide.

### Certification Regarding Lobbying

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

### Certification for Contracts, Grants, Loans and Cooperative Agreements

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

- (1) No Federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any Federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.
- (2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions.
- (3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

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

### Certification Regarding Nondiscrimination

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) is providing the Certification Regarding Nondiscrimination contained in Exhibit II-6 of the Grant Proposal Guide.

### Certification Regarding Flood Hazard Insurance

Two sections of the National Flood Insurance Act of 1968 (42 USC §4012a and §4106) bar Federal agencies from giving financial assistance for acquisition or construction purposes in any area identified by the Federal Emergency Management Agency (FEMA) as having special flood hazards unless the:

- (1) community in which that area is located participates in the national flood insurance program; and
- (2) building (and any related equipment) is covered by adequate flood insurance.

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) or Individual Applicant located in FEMA-designated special flood hazard areas is certifying that adequate flood insurance has been or will be obtained in the following situations:

- (1) for NSF grants for the construction of a building or facility, regardless of the dollar amount of the grant; and
- (2) for other NSF grants when more than \$25,000 has been budgeted in the proposal for repair, alteration or improvement (construction) of a building or facility.

### Certification Regarding Responsible Conduct of Research (RCR)

**(This certification is not applicable to proposals for conferences, symposia, and workshops.)**

By electronically signing the Certification Pages, the Authorized Organizational Representative is certifying that, in accordance with the NSF Proposal & Award Policies & Procedures Guide, Part II, Award & Administration Guide (AAG) Chapter IV.B., the institution has a plan in place to provide appropriate training and oversight in the responsible and ethical conduct of research to undergraduates, graduate students and postdoctoral researchers who will be supported by NSF to conduct research. The AOR shall require that the language of this certification be included in any award documents for all subawards at all tiers.

**CERTIFICATION PAGE - CONTINUED**

**Certification Regarding Organizational Support**

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) is certifying that there is organizational support for the proposal as required by Section 526 of the America COMPETES Reauthorization Act of 2010. This support extends to the portion of the proposal developed to satisfy the Broader Impacts Review Criterion as well as the Intellectual Merit Review Criterion, and any additional review criteria specified in the solicitation. Organizational support will be made available, as described in the proposal, in order to address the broader impacts and intellectual merit activities to be undertaken.

**Certification Regarding Federal Tax Obligations**

When the proposal exceeds \$5,000,000, the Authorized Organizational Representative (or equivalent) is required to complete the following certification regarding Federal tax obligations. By electronically signing the Certification pages, the Authorized Organizational Representative is certifying that, to the best of their knowledge and belief, the proposing organization:

- (1) has filed all Federal tax returns required during the three years preceding this certification;
- (2) has not been convicted of a criminal offense under the Internal Revenue Code of 1986; and
- (3) has not, more than 90 days prior to this certification, been notified of any unpaid Federal tax assessment for which the liability remains unsatisfied, unless the assessment is the subject of an installment agreement or offer in compromise that has been approved by the Internal Revenue Service and is not in default, or the assessment is the subject of a non-frivolous administrative or judicial proceeding.

**Certification Regarding Unpaid Federal Tax Liability**

When the proposing organization is a corporation, the Authorized Organizational Representative (or equivalent) is required to complete the following certification regarding Federal Tax Liability:

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) is certifying that the corporation has no unpaid Federal tax liability that has been assessed, for which all judicial and administrative remedies have been exhausted or lapsed, and that is not being paid in a timely manner pursuant to an agreement with the authority responsible for collecting the tax liability.

**Certification Regarding Criminal Convictions**

When the proposing organization is a corporation, the Authorized Organizational Representative (or equivalent) is required to complete the following certification regarding Criminal Convictions:

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) is certifying that the corporation has not been convicted of a felony criminal violation under any Federal law within the 24 months preceding the date on which the certification is signed.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE		SIGNATURE		DATE
NAME <b>Mary Gerrow</b>		<b>Electronic Signature</b>		<b>Apr 9 2014 1:33PM</b>
TELEPHONE NUMBER <b>520-626-6433</b>	EMAIL ADDRESS <b>maryg@u.arizona.edu</b>		FAX NUMBER <b>520-626-4130</b>	

## PROJECT SUMMARY

---

### **Overview:**

The tragic events of September 11th and the subsequent international conflicts had drastic effects on many aspects of society over the past decade. The many unintended and illicit uses of the Internet have also created significant cybersecurity challenges. Academics in the fields of computational science, information science, social sciences, engineering, and many other areas have been called upon to help enhance the government's and society's ability to fight terrorism, cybercrimes, and other security concerns. Facing the critical missions of national security and various data and technical challenges, many CISE researchers believe that there is a pressing need to develop the science of "Intelligence and Security Informatics" (ISI), with its main objective being the "development of advanced information technologies, systems, algorithms, and databases for security-related applications, through an integrated technological, organizational, and policy-based approach." Since 2001, CISE researchers have begun to develop significant, high-impact, and multi-disciplinary security-related research. A community of ISI researchers has emerged since the inaugural ISI Workshop sponsored by NSF, DHS, and DOD in 2003. The ISI series of meetings has since become the IEEE ISI Conference and has spun off the Pacific-Asia ISI workshop (PAISI), the ACM ISI-KDD Workshop, and the European ISI Conference (EISIC). The international ISI community consists of more than 1,500 scholars now, with about 70% in CISE disciplines (the rest in political and social sciences, international relations, criminal justice, etc.). Despite the importance of security informatics research for society and the progress made by the international ISI community, ISI CISE researchers are facing some pressing challenges, including: unique security characteristics, research data and testbed availability, need for collection, analytics and visualization tools, and challenges in evaluation, education and outreach. This DIBBs project will leverage many of the techniques and methods (if not the purpose, content, coverage and depth) developed in the course of our highly successful NSF CRI-funded(Computational Research Infrastructure) Dark Web portal project, used by more than 700 researchers in academia, governments and industry and adopted in major ISI and KDD workshops and conferences and for Ph.D. and Master's theses research.

This project is intended to create the Data Infrastructure Building Blocks (DIBBs) for the ISI community (thus DIBBs-ISI), a large archive and scalable research cyberinfrastructure for use by CISE researchers as well as related social and political scientists, studying a wide range of security related computational challenges. The research testbed and analytical tools will be accessible through a DIBBs-ISI portal interface that supports searching and downloading, and will provide a variety of open source and interoperable computational tools to support relevant analysis, such as social network analysis, authorship analysis, and data visualization. Opportunities for input and evaluation from these communities have been built in throughout the project, including an international Advisory Board of distinguished ISI scholars and outreach activities to major ISI conferences and the NSF SaTC (Secure and Trustworthy Cyberspace) and SFS (Scholarship-for-Service) cybersecurity research and education programs.

### **Intellectual Merit :**

The proposed DIBBs-ISI infrastructure will help collect and develop a suite of high-quality open-source interoperable security-related data and tools, made available through a scalable and flexible cyberinfrastructure, to support the critically important and growing international ISI research community. This DIBBs pilot demonstration project will significantly increase the breadth and depth of the previously CRI-funded Dark Web project and enhance the scalability and flexibility of a community research infrastructure. Important and complex security-related computational challenges can be addressed with the help of the proposed cyberinfrastructure.

### **Broader Impacts :**

The proposed project will help train many new graduate students in security informatics research and enhance offerings in security-related courses and training in the broad ISI community. In addition to supporting researchers in information, computer and social sciences, this project will also have significant utility for the national security sector, including the defense and intelligence communities, although that is not its primary purpose.

## TABLE OF CONTENTS

For font size and page formatting specifications, see GPG section II.B.2.

	<b>Total No. of Pages</b>	<b>Page No.* (Optional)*</b>
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)	1	_____
Table of Contents	1	_____
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) <b>(Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)</b>	15	_____
References Cited	3	_____
Biographical Sketches (Not to exceed 2 pages each)	14	_____
Budget (Plus up to 3 pages of budget justification)	28	_____
Current and Pending Support	9	_____
Facilities, Equipment and Other Resources	2	_____
Special Information/Supplementary Documents (Data Management Plan, Mentoring Plan and Other Supplementary Documents)	9	_____
Appendix (List below.) <b>(Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)</b>	_____	_____
Appendix Items:		

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



# CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research and Community (Pilot Demonstration Project)

## 1. Introduction

The tragic events of September 11<sup>th</sup> and the subsequent international conflicts had drastic effects on many aspects of society over the past decade. The many unintended and illicit uses of the Internet have also created significant cybersecurity challenges. Academics in the fields of computational science, information science, social sciences, engineering, and many other areas have been called upon to help enhance the government's and the society's ability to fight terrorism, cybercrimes, and other security concerns. Six critical mission areas have been identified where the CISE community can contribute, as suggested in the National Strategy for Homeland Security reports (2002, 2007), including: *intelligence and warning, border and transportation security, domestic counter-terrorism, protecting critical infrastructure, defending against catastrophic terrorism, and emergency preparedness and response*. Facing the critical missions of national security and various data and technical challenges, many CISE researchers believe that there is a pressing need to develop the science of "Intelligence and Security Informatics" (ISI) (Chen 2006), with its main objective being the "development of advanced information technologies, systems, algorithms, and databases for security-related applications, through an integrated technological, organizational, and policy-based approach."

Within ISI, KDD (Knowledge Discovery from Databases) techniques can play a central role in improving counter-terrorism and crime-fighting capabilities of intelligence and security agencies by reducing cognitive and information overload. KDD refers to non-trivial extraction of implicit, previously unknown, and potentially useful knowledge from data. KDD techniques promise easy, convenient, and practical exploration of very large collections of data for organizations and users, and have been applied in marketing, finance, manufacturing, biology, and many other domains (Fayyad and Uthurusamy, 2002). Many of the KDD technologies could be applied in ISI studies. Keeping in mind the special characteristics of crimes and security-related data, we categorize existing ISI technologies into six classes: *information sharing and collaboration, crime association, text mining, spatial and temporal pattern mining, and criminal network analysis*. In Figure 1 we present the KDD framework for ISI research, with the horizontal axis being the security concerns and the vertical axis being the six classes of data mining techniques (Chen 2006).

### 1.1 Intelligence and Security Informatics (ISI) Research and Community

Shortly after 2001, many CISE researchers have begun to develop significant, high-impact, and multi-disciplinary security-related research. A community of ISI researchers has emerged since the inaugural ISI Workshop sponsored by NSF, DHS, and DOD in 2003. The ISI series of meetings has since become the IEEE ISI Conference and has spun off the Pacific-Asia ISI workshop (PAISI), the ACM ISI-KDD Workshop, and the European ISI Conference (EISIC). The international ISI community consists of more than 1,500 scholars now, with about 70% in CISE disciplines (the rest in political and social sciences, international relations, criminal justice, etc.). As explained in the web site of the most recent IEEE ISI Conference, held in Seattle, Washington on June 4-7, 2013 (THEME: Big Data, Emergent Threats and Decision-Making in Security Informatics; <http://isiconference2013.org/pgs/>):

"Intelligence and Security Informatics (ISI) research is an interdisciplinary research field involving academic researchers in information technologies, computer science, public policy, bioinformatics, and social and behavior studies as well as local, state, and federal law enforcement and intelligence experts, and information technology industry consultants and practitioners to support counterterrorism and homeland

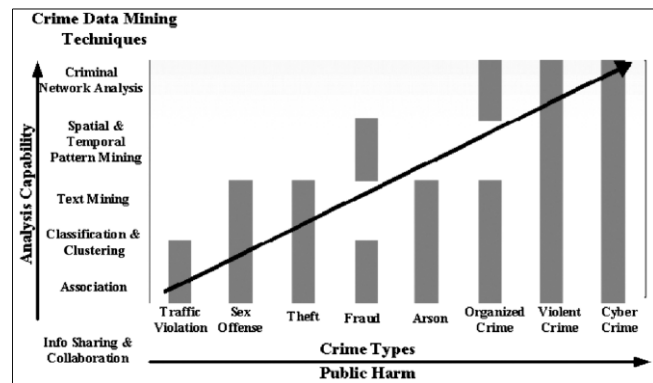


Figure 1. A KDD (knowledge discovery from databases) research framework for ISI

security missions of anticipation, interdiction, prevention, preparedness and response to terrorist acts. The annual IEEE International ISI Conference series was started in 2003. Meetings have been held in Tucson, AZ (twice); Atlanta, GA; San Diego, CA; New Brunswick, NJ; Taipei, Taiwan; Dallas, TX; Vancouver, Canada; and Beijing, China, and Washington D.C. Proceedings of these ISI meetings and workshops have been published by the IEEE Press and in the Springer Lecture Notes in Computer Science (LNCS) series.” Some of the sample IEEE ISI 2013 conference topics include:

**(1) Big Data in Security Informatics**

- Computer or cybercrime investigations and digital forensics
- Criminal data mining and network analysis
- Criminal/intelligence information sharing and visualization
- Crime pattern recognition and modeling tools
- Web-based intelligence monitoring and analysis
- Deception and intent detection

**(2) Emergent Threats**

- Cyber-infrastructure design and protection
- Bio-terrorism tracking, alerting, and analysis
- Emergency response and management
- Computer forensics and crime lead discovery
- Terrorism Informatics
- Terrorism related analytical methodologies and software tools

**(3) Decision-Making in Security Informatics**

- Fraud detection
- Cybercrime and social impacts
- Corporate going concerns and risks
- Corporate governance and monitoring
- Board activism and influence

**1.2 ISI Challenges and Community Needs**

Despite the importance of security informatics research for society and the progress made by the international ISI community, ISI CISE researchers are facing some pressing challenges:

- **Unique Security Characteristics:** Most of the ISI CISE researchers are well trained in computational and information sciences, including data mining, machine learning, social network analysis, machine learning, computational linguistics, cybersecurity, HCI, etc. However ISI research exhibits unique problems that are not present in traditional academic research, e.g., adversarial data mining with shifting opponents (criminals), deception and intent detection in criminal communications (social media propaganda), multilingual challenges in difficult and rare languages (Farsi, Urdu), etc. ISI researchers often need to re-examine the underlying assumptions with the security phenomena of interest.
- **Research Data and Testbed Availability:** Unlike traditional CISE research which can rely on many open-source and publically available data (e.g., UC Irvine machine learning data, Netflix movie recommendation testbed, i2b2 clinical discharge data), ISI research often demands unique security-related testbeds or contents. Although Open-Source Intelligence has become critical for conducting ISI research, many such contents have only begun to accumulate in scale and quality, e.g., the Dark Web Portal (details to follow). More needs to be done to foster active ISI testbed development, information sharing, and system interoperability.
- **Needs for Collection, Analytics and Visualization Tools:** In addition to security data sharing, ISI researchers will also benefit from sharing critical open-source collection, analytics and visualization tools in the community. As most of the security problems cannot be resolved with one single technique or approach, ISI researchers will benefit from computational tools that can convert the raw secure data into proper representations for their own targeted analysis, e.g., social network analysis researchers will benefit from forum postings that are tagged with topics and sentiments, HCI scholars will benefit from foreign texts that have been machine translated and displayed in visual summaries, etc.
- **Evaluation, Education and Outreach:** The ISI community testbeds and tools will need to be individually evaluated and collectively assessed by the community. A DIBBs infrastructure and portal will provide a vehicle for community sharing and feedback. Outreach into the international IEEE ISI, EISIC, PAISI, and ISI-KDD communities of CISE researchers will help broaden the reach and impact of this critically important emerging research field. In addition, the DIBBs infrastructure can also provide an advanced testbed for helping to develop a new generation of security professionals through

the highly successful and growing NSF Cybersecurity SaTC (Secure and Trustworthy Cyberspace) and SFS (Scholarship-for-Service) programs, with which our project team (Drs. Chen, Yang, and Thuraisingham) has significant existing involvement (more details to follow).

## 2. Previous CRI Research: The Dark Web Portal

The challenges in collecting and managing large-scale longitudinal open-source security-related contents and analytical tools from various international data sources are numerous. However, our proposed DIBBs pilot demonstration project will leverage many of the techniques and methods (if not the purpose, content, coverage and depth) developed in the course of our highly successful previously NSF CRI (Computational Research Infrastructure) funded Dark Web portal project (Chen et al., 2011b; Chen, 2012).

### 2.1 Dark Web Portal Overview

Dark Web project is an internationally recognized long-term terrorism informatics research program that examines international terrorism and extremism via a computational, data-centric approach (Chen, 2012). It has been funded in part by the National Science Foundation and the Defense Threat Reduction Agency. We collect web content generated by international terrorist groups, including web sites, forums, chat rooms, blogs, social networking sites, videos, virtual worlds, etc. (Chen, et al., 2004; Zhou et al., 2006; Fu et al., 2010; Chen et al., 2011b; Chen, 2012). We have also developed various multilingual data mining, text mining, and web mining techniques to perform link analysis (Zhou, et al., 2005), sentiment analysis (Abbasi et al., 2008) authorship analysis (Abbasi & Chen, 2005; 2008), and video analysis (Huang et al., 2010) in our research. Our work has resulted in part in the Dark Web portal, which contains over 29,000,000 forum threads/messages in English, Arabic, French, German, and Russian, with detailed metadata and raw html contents (Zhang et al., 2010b; Chen et al., 2011b).

### 2.2 Dark Web Success and Lessons Learned

With the funding support from the NSF CRI program (2007-2012), we have developed a collection, analytics, and web portal infrastructure for accessing, downloading and analyzing selected Dark Web contents. Since its operation in 2008, the Dark Web portal offers the ISI community a unique research testbed for their CISE research. Researchers, analysts, and others wishing to use the portal request an account. The account request form records the researcher's name, contact information, affiliation and organization type, and reasons for using the portal. Users who have logged into the portal can browse the forums, search for keywords across all forums, and then translate, print, and download their results. Users may input their search terms in English or any other language used in the portal; search terms are automatically translated to the other portal languages so that matching results can be returned from all forums. A sample screen is shown in Figure 2.

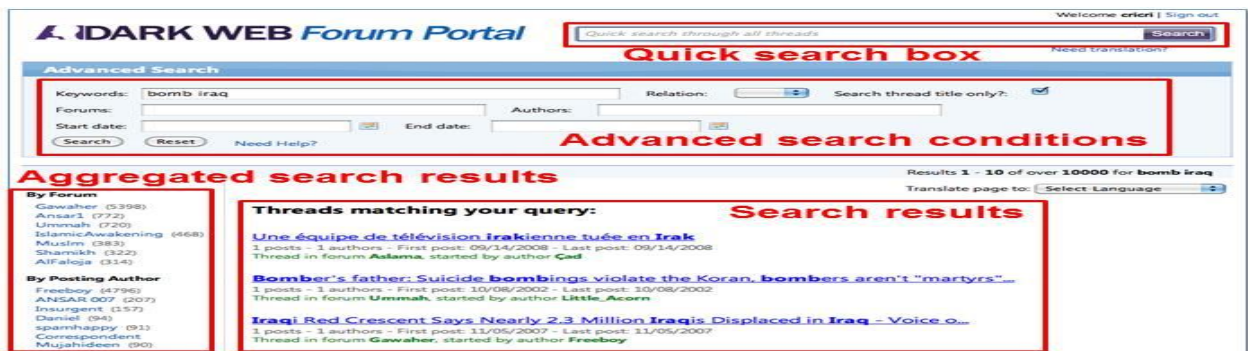


Figure 2. Dark Web Portal Search Interface

Usage of the portal grew over the life of the project despite the project ending in 2012. The majority of the 700+ registered users are from the United States, but in all, over 20 countries are represented, with the larger numbers of users coming from Australia, Canada, China, Denmark, France, Germany, Italy, India, the Netherlands, and the UK. Users come from academia, government (including local law enforcement, military, and intelligence units), industry, and non-profits. We continue to maintain the portal and keep it

running, though new data has not been added since the project ended. A summary of user statistics is shown below:

- Number of portal users requesting access by year (in parenthesis): 2008 (29), 2009 (162), 2010 (207), 2011 (75), 2012 (155), 2013 (93). Most users learned about the portal from the IEEE ISI, EISIC, PAISI, and ISI-KDD events, others through Google searches, our project web site, or referrals.
- User statistics by affiliation (number of users in parenthesis): academic (420), government (153), industry (85), law enforcement (15), non-profit (29), and other (34). Among the academic, most of the users belong to either computational/information science or social/political science disciplines.
- There are also acknowledgements for using the Dark Web Portal for Ph.D. or Master's theses research, e.g., "Media Use in Conversion into Islamic Radicalism: Why This Person and Not That Person, Why Here, Why Now" by Frank Hairgrove, University of Wisconsin-Madison, 2011, Ph.D. thesis; "Radicalization Online: Patterns of Social Interaction on the Al-Faloja and As-Ansar Forums," by Lucas Lenselink, Utrecht University, 2011, Master's thesis.

Reasons for using the Dark Web Forum Portal are as varied as the users themselves. University students in information science, computer science, social sciences, and intelligence analysis have used it for both coursework and theses. Scientists in computer science, systems engineering, and related disciplines have used it, for example, to research non-kinematic target tracking, to study the use of Bayesian networks for conditional probability, and to develop text mining techniques for authorship and content analysis. In our many formal and informal interactions with ISI researchers at workshops and meetings, we have received many positive and encouraging comments regarding the value of such a community-wide resource.

### **2.3 ACM ISI-KDD Workshop and Challenge**

In 2009 and 2010, with NSF CRI funding, we (Drs. Yang and Chen) organized two ISI workshops in conjunction with the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (the largest international data mining conference, with typically more than 700 participants). The first workshop was the ACM SIGKDD Workshop on CyberSecurity and Intelligence (CSI-KDD 2009) and the second was the ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD 2010). In both events, the focus was the study of the development and use of advanced information technologies and systems for security-related applications. In addition, we organized the "ISI-KDD Challenge" in the ISI-KDD workshop, where participants are challenged to "find the more radical and infectious threads, members, postings, ideas and ideologies" in the Dark Web Forum Portal from within several complete multi-year extremist forums. Participants of the workshop are granted access to more than 1,000,000 messages and postings by these radical members. Participants are then encouraged to develop novel computational techniques and algorithms, e.g., linguistic analysis, topic extraction, multilingual text parsing, sentiment analysis, social network analysis, time-series analysis, etc. for the challenge. Participants are requested to describe their methods and results in papers.

In CSI-KDD 2009, eight papers were selected for presentation in the workshop and printed in the proceedings. In ISI-KDD 2010, ten papers were selected for presentations; among them four papers were developed based on the Dark Web Portal contents. Sample accepted papers in ISI-KDD 2010 include:

- "Two-stage Approach for Unbalanced Classification with Time-varying Decision Boundary: Application to Marine Container Inspection," by Richard Hoshino, Wayne Oldford and Mu Zhu
- "Fuzzy Association Rule Mining for Community Crime Pattern Discovery," by Anna Buczak and Christopher Gifford
- "Applying Interestingness Measures to Ansar Forum Texts," by David Skillicorn (ISI-KDD Challenge paper)
- "Topic-Based Social Network Analysis for Virtual Communities of Interests in the Dark Web," by Gaston LHuillier, Hector Alvarez, Sebastian Rios and Felipe Aguilera (ISI-KDD Challenge paper)

Details about ISI-KDD 2010 can be found in Yang & Chen (2010) published in *SIGKDD Explorations*. Since the expiration of the NSF CRI funding in 2012 we have continued to support the Dark Web portal and organized ISI-KDD 2012 (held in Beijing, China), with seven presentations and two ISI-KDD Challenge papers.

## 2.4 From Dark Web Portal to DIBBs-ISI: Developing an ISI Community Infrastructure

Despite the success of the Dark Web CRI project and the overwhelming support in the ISI community, we have identified several areas for improvement to and enhancement of this DIBBs-ISI project:

- **ISI Collection Breadth and Depth:** Suggestions have been made in several of the ISI-KDD and IEEE ISI meetings regarding inclusion of other security-related testbeds, ranging from cybersecurity data to online phishing web sites. Several major ISI research groups have expressed their interest in contributing their contents and to help develop an ISI-community infrastructure. Some of these contents will go beyond unstructured multilingual social media communications to include videos, image attachments, hacker source code and malware, etc. Each of these content types will require careful database design and harvesting.
- **Analytical, Visualization and Access Support:** In association with the expanded ISI content coverage, there will be a need for developing and sharing open-source tools developed by researchers in the community. Some of the tools can be collected and made available for downloads; others may need to be integrated into a portal system to assist in more targeted security analysis and exploration. Our DIBBs-ISI research team will lead the effort in developing our collection and analytical tools. We will also use our developments to help identify future infrastructure and community needs.
- **Scalable, Flexible and Interoperable Community Infrastructure:** One of the most critical and novel enhancements in our proposed DIBBs-ISI project is to develop an infrastructure for other ISI researchers to contribute their contents and tools. A flexible Collection Manager and Testbed APIs will need to be developed to support scalable and interoperable data ingest and content management. In addition, we also plan to investigate the development of a community-wide collaboration environment for linking different research groups (e.g., content providers with analytical groups) and for sharing other community news and resources. A DIBBs-ISI Advisory Board will be created to help guide the development of community policy and outreach activities (more to follow below).

## 3. DIBBs-ISI Development Framework, Testbed, and Infrastructure

In the following subsections, we describe the proposed DIBBs-ISI architecture and the physical infrastructure. Comparing to the Dark Web CRI project, the proposed DIBBs-ISI infrastructure will be more flexible and scalable in meeting the needs of a broad ISI community.

### 3.1 System Architecture and Components

Architecturally, the DIBBs-ISI infrastructure will consist of three loosely coupled components: *Testbed Builder*, *Testbed Management Framework*, and *DIBBs-ISI Portal*. The communication between these components will be implemented as WebServices to ensure cross-platform compatibility and scalability. The development of these components in this way allows us to leverage our previous Dark Web CRI work and other open source projects. Figure 3 illustrates all components and shows the main data flows between them and the underlying data storage.

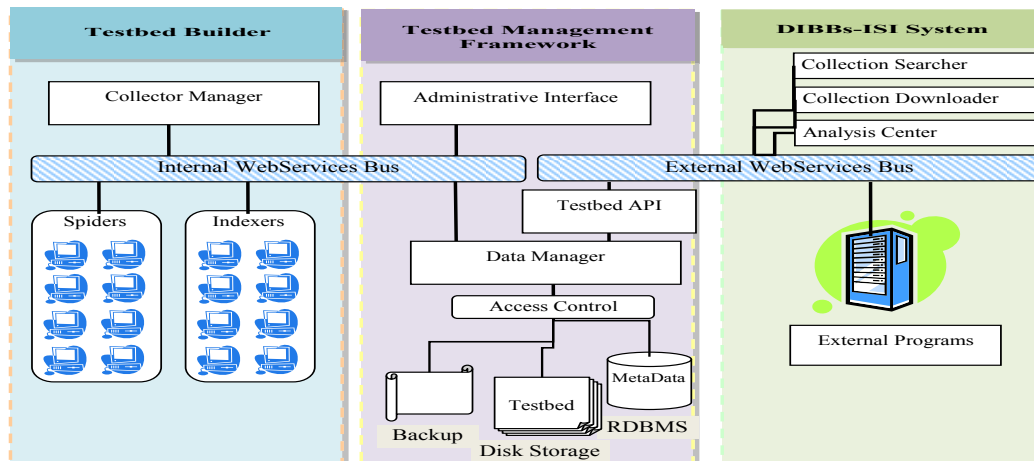


Figure 3. DIBBs-ISI System Architecture

### **Testbed Builder - Collection Building**

Developing a complete DIBBs-ISI testbed requires tremendous effort in continuously collecting open-source security contents from the Internet. This poses a non-trivial challenge to the underlying infrastructure especially for network bandwidth, storage disks, and software scalability. Web spiders are programs that implement Hypertext Transfer Protocol (HTTP) to browse the targeted Web sites in an automated manner (Albertsen, 2003). To develop the DIBBs-ISI testbed, the spider software must have multi-lingual and multi-threading capability. We will improve our previous Dark Web spiders, which are cookie-aware for accessing password-required pages and scalable for up to one million Web content items. The improved DIBBs-ISI spiders will be spread across several virtual machines and will run multiple instances of the spiders. The Web page indexers will extract textual terms from given Web pages and build a table indicating where the terms are used. From our observation, groups and organizations utilize not just HTML but also multimedia content including audio and video files. Our indexer will be based on the Apache Lucene project, a high-performance, full-featured text search engine library written entirely in Java. The Collector Manager module is at the heart of the testbed builder. This module will run as a daemon and communicate with an array of spiders and indexers to moderate their workload and monitor status. It is also responsible for scheduling periodic collection updates and collect completed Web site collections for Data Manager.

### **Testbed Management Framework - System Management and Interoperability**

To automate the testbed building and serving processes, the testbed management framework is critical for bridging the gap between the testbed portal and builder. The management framework has three modules: the *Data Manager*, *Testbed API*, and *Administrative Interface*. The Data Manager module has three main functionalities: a) collection and metadata storage; b) testbed access control; and c) data backup scheduling. For faster searching, we will use an open source relational database system to manage the metadata of the testbed. Lastly, the data manager module will connect to a hundred-terabyte tape library for regular backup. The Testbed API is the main programming interface to the testbed data. For best interoperability, this set of APIs will be implemented as WebServices, leveraging HTTP and XML standards. This API will provide listing, searching, and retrieval of the testbed data. The Administrative Interface module implements a Web-based user interface for testbed administrators. This module provides a wide range of functionalities including testbed builder control, data management control, backup schedule management, user account management, and log management.

### **DIBBs-ISI Portal – Search and Analysis**

The DIBBs-ISI Portal provides the interface to the ISI researchers. The portal consists of: *Collection Searcher*, *Collection Downloader*, and *Analysis Center*. Other external programs can then be used to analyze the retrieved or downloaded contents. A sample search interface is shown in Figure 2 (based on the Dark Web Portal).

### **3.1 Physical Infrastructure**

In this section, we describe the physical infrastructure needed to support the proposed DIBBs-ISI work.

#### **Facilities**

The repository will be hosted by the MicroAge Lab, a research laboratory founded by the PI, Dr. Hsinchun Chen, in 2000. This lab is dedicated to providing support to education and research activities. The existing server room can readily accommodate the additional servers, storage units, and power supplies necessary for this proposal in terms of space, power, and cooling capacity. An additional benefit to using this resource is the ability to leverage the existing backup infrastructure which includes a large IBM robotic tape library and an installed backup controller/storage manager. The lab also has a dedicated fiber network link to the University's central network, which alleviates potential network conflicts when spidering and updating the data collection.

#### **Data Security and Networking**

Physical security for the installation will be strong. The MicroAge Lab is alarmed and uses a card-swipe that tracks all lab visitors upon entrance. Data network security will be managed by a dedicated firewall appliance in conjunction with server firewalls. All servers will run anti-virus software and all access will be restricted to secure channels such as secure shell (ssh), server socket layer (ssl), or virtual

private network (VPN). Data will be backed up to an IBM Ultra scalable Robotic Tape library using Tivoli (already in use, will leverage existing licenses) as the storage and backup controller. Tape copies will be maintained in the library and an archival data set will also be maintained at a secure offsite location.

The infrastructure will support two separate networks, a data network and a storage network. Both networks will use industry standard twisted pair gigabit Ethernet connections with redundant gigabit switches. The data network will be connected to the campus backbone through a dedicated firewall and then through an existing fiber link dedicated to the hosting lab which will connect it to central computing. The storage network will use Internet Small Computer System Interface (iSCSI) infrastructure instead of a fiber based Storage Area Network (SAN) infrastructure to ensure support for future expansion in a cost efficient manner.

#### **Management, Storage, and Application Infrastructure**

The management infrastructure will consist of a single machine that serves as a controller for the multiple spidering nodes, which will individually crawl separate URLs as assigned by the controller. The controller and spidering servers will be 4-socket 48-core servers with 512 GBs RAM and high duty cycle SCSI drives. The storage infrastructure will be based on an iSCSI storage area network providing access to two iSCSI gateways that each have a 96 TBs storage array attached. The dedicated iSCSI network will provide significantly better performance than a Network Attached Storage (NAS) architecture without incurring the overhead costs or tuning requirements of a dedicated fiber SAN. The proposed architecture is robust, redundant, and scalable with room for future growth including the capacity to attach additional storage arrays to either gateway as necessary.

#### **4. DIBBs-ISI Collection and Analytics Development Plan: GeoPolitical Web, Hacker Web, and AZPhish Web**

In the following subsections, we describe the several planned expansions of the proposed DIBBs-ISI collection and analytics development efforts. For the Geopolitical Web testbed, we will expand our previous Dark Web collection efforts to include the diverse social media contents prevalent in many volatile geopolitical regions. In Hacker Web, we aim to collect hacker forums, IRC, and botnet C&C contents and develop tools in support of hacker community analysis. The Hacker Web will leverage our recently funded NSF SaTC (Secure and Trustworthy Cyberspace) project, which will develop selected collection and analytical techniques for possible adoption by other ISI researchers. Lastly, we describe our AZPhish Web testbed that aims to facilitate data mining and cybersecurity research for anti-phishing. We hope to use these three testbeds to demonstrate the value of the DIBBs-ISI infrastructure and draw attention to our project. Other new ISI testbeds will be suggested by the DIBBs-ISI advisory board and contributed from the ISI community in the future.

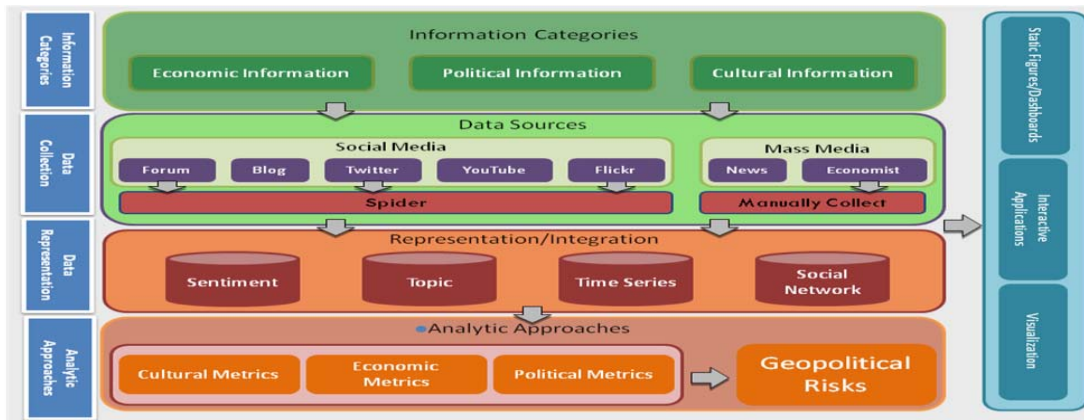
##### **4.1 The GeoPolitical Web Testbed**

Given the exponential growth of the Web 2.0, there is continuing opportunity to use social media to study and analyze country-based as well as non-state social movements, current opinions, and their dynamics. For example, the “Arab Awakening” events garnered significant attention from researchers in many disciplines (The Economist, 2011). The GeoPolitical Web research testbed will ultimately be comprised of data containing millions of multilingual social media contents (such as forum messages, tweets, videos, news, and comments) from volatile regions around the world (for example, Afghanistan, Iraq, Somalia, Yemen, etc.). The archive will be accessible through an interface that supports searching in meaningful ways, and will provide a variety of tools to support relevant analysis, such as social network analysis, authorship analysis, and data visualization. We will leverage a Defense Threat Reduction Agency-supported project that provides an initial data collection. The testbed will be made available through our proposed DIBBs-ISI infrastructure. The archive will be useful not only to computer and information scientists, but also to social and political scientists studying a wide range of computational and social problems.

##### **Development Framework**

The GeoPolitical Web development framework for collecting, representing, and analyzing relevant socio-cultural data sources is shown in Figure 4 (Chen et al., 2011a).





**Figure 4.** Development Framework of the Geopolitical Web

The information categories needed in GeoPolitical Web have been identified as economic, political, and cultural information. For the proposed work, we will leverage existing international news sources (as the baseline content) and expand on the availability of new social media contents. Local political, social and cultural information will be gathered primarily through social media such as forums, websites, blogs, video sharing sites, and tweets. The textual data will be processed by the different tools and techniques listed in the data representation category, such as sentiment analysis, topic extraction, time series representation, and social network analysis. Many of these computational techniques and tools have become mature for supporting analysis. Finally, using metrics and representations from the collected economic data and social media, context-relevant computational and social analyses can be performed for various geopolitical questions motivated by the issues and conditions in different parts of the world.

### Initial Collection

To conduct our pilot study, we selected a set of countries considered to be at risk, including Somalia and the Maghreb region of Africa (Morocco, Algeria, Tunisia, Libya, Mauritania); Afghanistan, Iraq, and Yemen of the Middle East; and Indonesia for Southeast Asia/Oceania (Chen et al., 2011a). These “at-risk” countries have been identified as such by sources such as *Euromoney* (2011) and the Failed States Index. The initial set of countries was also chosen to represent a diversity of languages, thus allowing deeper study of multi-lingual computational issues (Munson, 2001; Abbasi & Chen, 2005; Chen, 2012). English, Arabic and French, for example, are the three dominant languages used in social media relating to the Arabic Spring. Languages such as Pashto, Somali, and Indonesian have rarely been studied in computational linguistics but are critical for understanding those volatile regions. Table 1 provides a snapshot of the GeoPolitical Web data collection.

**Table 1.** Data Collected for the GeoPolitical Web Prototype

Medium	Scope	Quantity by Country		Time Span
Forums	<ul style="list-style-type: none"> <li>- Wide discussion on universal topics</li> <li>- Postings organized by threads</li> <li>- First number is number of forums collected; second number indicates number of messages</li> </ul>	AFG:	5 443,465	2002-2011
		EGY:	5 2,366,943	2000-2010
		IND:	5 1,731,394	2000-2012
		IRQ:	8 3,243,636	2002-2012
		LEB:	4 1,413,130	2005-2012
		PAK:	7 4,506,890	2003-2010
		SAU:	5 8,752,444	2001-2012
		SOM:	7 816,653	2007-2011
		TUN:	5 973,538	2000-2012
		YEM:	10 1,392,293	2002-2012
Twitter	<ul style="list-style-type: none"> <li>- “Tweets” – short postings</li> <li>- Numbers of tweets collected to date</li> </ul>	SOM:	41,958	2010-2011
		YEM:	79,827	2010-2011
Country abbreviations:	AFG, Afghanistan	IRQ: Iraq	PAK, Pakistan	TUN, Tunisia
	EGY, Egypt	LEB, Lebanon	SAU, Saudi Arabia	YEM, Yemen
	IND, Indonesia	MAG, Maghreb region	SOM, Somalia	



Forum contents are longitudinal, covering from 2000 to 2012, and diverse in subject discussions. We have collected more than 25,000,000 messages from 60 international forums to date. Tweets can only be collected via the Twitter API for new content (i.e., no back archive available for spidering) and thus can cover only recent events. We have only begun to collect selected twitter contents about Yemen and Somalia. International and local news can be spidered automatically. After being collected, all forums postings, tweets, and news articles are parsed into a centralized database.

#### 4.2 The Hacker Web Testbed

Our recently funded NSF SaTC project entitled “Securing Cyber Space: Understanding the Cyber Attackers and Attacks via Social Media Analytics” (the Hacker Web project) offers another opportunity for developing a testbed and tools for use at DIBBs-ISI. While basic research will be conducted within the SaTC project, we anticipate that the valuable hacker community contents we will collect and analytical tools we will develop could be enhanced for inclusion in our proposed DIBBs-ISI project. The Hacker Web project aims to develop an integrated and scalable computational social media collection and analytics framework in support of the cyber attacker community analysis. Our multi-disciplinary research team of computational and social scientists (Drs. Hsinchun Chen, Salim Hariri, Tom Holt, Ron Breiger) will address important research questions of relevance to hacker skills, community structure and ecosystem, contents and artifacts, and cultural differences.

#### Development Framework

The proposed Hacker Web testbed development framework is shown in Figure 5.

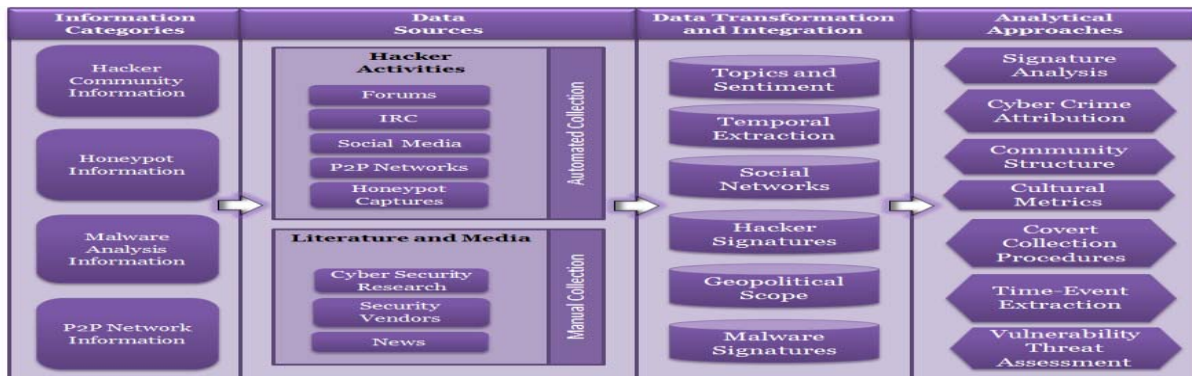


Figure 5. Development Framework of the Hacker Web

We start by identifying several important categories of information necessary for cyber security investigation. Our research will focus on hacker community information (the actors) and honeypot information (malware output), to be supplemented by further malware analysis and selected emerging P2P network information. Then, data sources for each information category are identified and collected to assist in our hacker community analysis. We plan to develop automated techniques for collecting major U.S., Russian and Chinese hacker forums and IRC contents. We will also explore additional social media, P2P networks, and honeypot captures. In addition, manual collection methods will be deployed for emerging cyber security research and news and other security vectors based on our social science and security analysis research questions. We will leverage our extensive experience in social media analytics for from our NSF funded Dark Web research for topics and sentiment, temporal extraction, and social networks. Additional hacker and malware signatures (e.g., programming languages used, attack targets, source code used) and other geopolitical information (e.g., locations) will be identified to assist in hacker community analysis. Lastly, numerous types of social science and security analyses will allow us to gain new perspectives and knowledge from the acquired data: hacker signature analysis (profile), cyber crime attribution (linking malware to actors), hacker community structure (and skills), and cultural metrics identification (for US, Russian, and Chinese groups). In addition, our research will help with time-event extraction, covert hacker community content collection, and vulnerability threat assessment.

#### Hacker Forums Collection

We believe much of our Dark Web forum spidering technology can be adopted for hacker forums collection. Dr. Tom Holt, who is one of the leading experts in hacker community research (Holt 2007; Holt

& Kilger 2012), will lead the effort in identifying key U.S., Russian and Chinese hacker forums in various web sites and public ISPs. The AI Lab, headed by Dr. Chen, will lead the effort in adapting Dark Web spidering tools for hacker community forums. Careful spidering and proxy setting will be developed to avoid detection and bypass anti-crawling mechanisms. Forum contents collected will be stored in a local cached relational database for research purposes. Selected malware source code and attachments will be collected and made available through DIBBs-ISI.

### **Honeypots and IRC Channels Collection**

In addition to hacker forums collection, there is also value in applying data collected through honeypot technologies based on their ability to mimic an unpatched vulnerable end user computer. Honeypots are systems that are configured to simulate computer environments with software vulnerabilities; the idea is to have wild malware exploit honeypot vulnerabilities so that the malware can be captured and studied in a sandboxed environment. All code execution, system changes, and network traffic are tracked and logged within a honeypot, letting security researchers understand the nature of some particular malware (Mielke & Chen, 2008; Zhu et al., 2008). For this project we propose combining the analysis of forum data with analyses of active malware acquired through the use of various collection devices. Once acquired, we will then infect honeypot systems managed by the research team and analyze the traffic and behavior of malware through the use of honeywall software (The HoneyNet Project, 2003). Honeypot approaches towards C&C identification will be implemented. By observing outbound network connections attempted by captured malware, researchers may potentially reveal botnet C&C channels and other hacker-related web addresses. We anticipate great value in including the Hacker Web testbed and tools in our proposed DIBBs-ISI infrastructure.

### **4.3 The AZPhish Web Testbed**

Phishing websites are fraudulent websites used to deceive unsuspecting Internet users (Abbasi and Chen 2009). Phishing websites have become increasingly pervasive, generating billions of dollars in fraudulent revenue, disseminating malware, and constituting a major form of insider threat (Dhamija et al. 2006; Hong 2012). Verified phishing incidents reported to PhishTank went up by 31% between 2011 and 2012 (PhishTank 2013). According to a Gartner report on industry adoption of security tools, increased phishing attacks have caused demand for web fraud detection software to reach all-time highs (Gartner 2011). Phishing websites span numerous domains, including financial, medical, legal, retail, social networking, and search/portal websites, just to name a few. The authentic and legitimate appearance of phishing websites makes it difficult for users to identify them as fraudulent (Dhamija et al. 2006; Jagatic et al. 2007). Furthermore, the performance of existing anti-phishing methods is not always up-to-par (Abbasi and Chen 2009b; Hong 2012). Consequently construction of large-scale phishing website testbeds is important to facilitate data mining research on predictive analytics for anti-phishing (Abbasi and Chen 2009b), as well as behavioral research to better understand user susceptibility to phishing website-based attacks (Abbasi et al. 2012).

### **Development Framework**

To address these issues, we propose the following AZPhish Web system described in Figure 6. The system monitors appropriate data sources in real-time, downloads content to a quarantined storage location, and then carefully extracts key text, source code, image, and linkage features and transfers them to a structured database, along with necessary raw content. The collection process leverages our extensive prior experience collecting adversarial content from the web (Fu et al. 2010). For search and social media sources, where the phishing status of a given URL is not immediately clear, we leverage our extensive prior experience with phishing website detection to identify the subset of candidate URLs as phish (Abbasi and Chen 2009a; 2009b; Abbasi et al. 2010; Abbasi et al. 2012a).

### **Initial Collection**

Using the proposed AZPhish Web framework, we have presently collected a large quantity of concocted and spoof phishing websites pertaining to various industry sectors. For each website, the collection includes all web pages, images, in/out links, source code/scripts, etc. In total, the current collection encompasses over 1.2 million phishing web pages and 2.7 million images. As part of the proposed research, we intend to collect thousands of additional concocted and spoof websites, and make the phishing website database publicly available to the ISI and CISE research community. We also plan on making source code for the AZPhish Web tool publicly available.

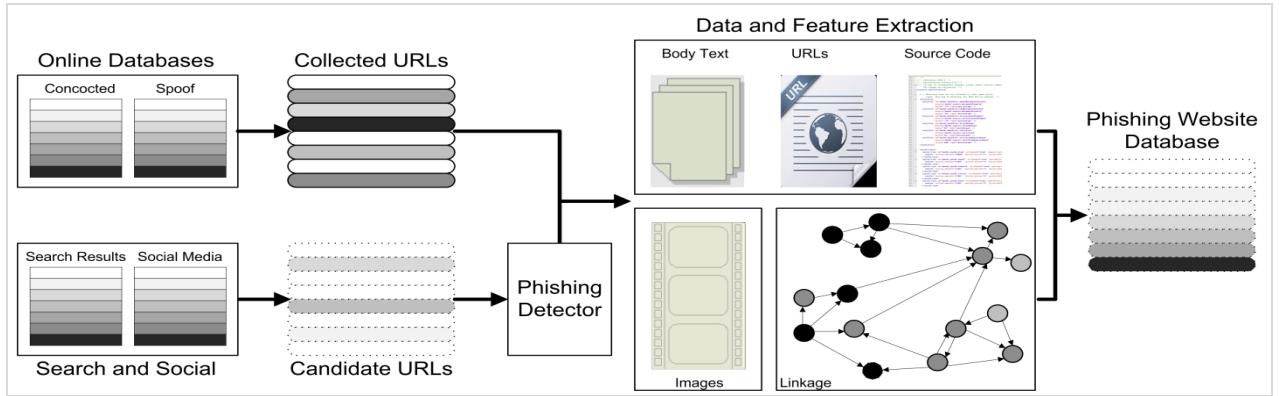


Figure 6. Development Framework of the AZPhish Web

## 5 User Involvement and Evaluation Plan

### 5.1 DIBBs-ISI Advisory Board

User acceptance is critical to the success of the development of the DIBBs-ISI infrastructure and, once it has been built, to the continuation of its existence. We have therefore planned for an integrated user-centered design process and taken care to build user input into each stage of the development process. We have successful interface design experience from previous projects to draw on, and will also work with our HCI expert to help design the best methods for leading this process. Toward this end, we have begun composing an advisory board with members from a variety of ISI-related CISE disciplines. Most of these members are distinguished researchers who have significant ISI research and development experience. Several of them have been involved in the past IEEE ISI, EISIC, PAISI, ISI-KDD, and NSF SaTC/SFS activities. All members will provide input and suggestions on policy, data selection, the analytical tools to be developed, and the portal functionalities.

We will interact with advisory board members regularly and in a variety of ways. Once the project is underway, we will have a “kick-off” advisory board meeting using virtual meeting or other meeting software. We will introduce members to each other, introduce the project in more detail, and follow with members discussing their expected uses of the portal. After the kick-off meeting, we will interact regularly with board members by sending bimonthly updates which will include opportunities for them to provide feedback related to current and near-term tasks. We will also send out an annual survey to help with evaluating program outcomes and ensure we are on track. We also regularly attend some of the same ISI series of conferences that our advisory board members attend. As we did with Dark Web, we will use those opportunities to schedule in-person meetings with them, and also to disseminate project information to additional potential users. When more intensive feedback is needed (for example, in evaluating wireframes for the portal interface, or in testing analytical tools), members will be solicited to invite additional colleagues and students to participate in more formal evaluations, in addition to our evaluations using local audiences. Members who have agreed to serve to date and who plan to use the portal for their research interests are listed in Table 2; we will continue to solicit new members as the project progresses.

Table 2. Advisory Board Members to Date

Name	Affiliation	Research areas
Michael Chau	University of Hong Kong, Hong Kong (past PAISI chair)	Modelling social media infection
Uwe Glaesser	Simon Fraser University, Canada (past IEEE ISI chair)	Computational criminology modelling and analysis
Ee-Peng Lim	Singapore Management University	Information extraction and text mining; network linking
Lisa Kaati	Uppsala University, Sweden (past EISIC chair)	Terrorism analytics, social simulation
Salim Hariri	University of Arizona (NSF SaTC and SFS project)	Cyber security, autonomic computing
David Skillicorn	Queen’s University, Canada (past SIAM Data Mining Conf. chair)	Adversarial knowledge discovery, security inductive modelling, social network disruption

Alan G. Wang	Virginia Tech (past PAISI chair; IEEE ISI 2014 chair)	Deception detection, Bayesian networks modelling
Uffe Kock Wiil	U of South Denmark, Denmark (past EISIC chair)	Dark network analysis, software engineering, knowledge management
Jennifer Xu	Bentley University	Dynamic social network analysis and visualization.
Dongsong Zhang	U of Maryland at Baltimore County	Mobile computing, security analytics
Beichuan Zhang	University of Arizona	Network security, internet protocols

## 5.2 Evaluation Plan

Our user-centered design process includes a system evaluation framework that is iterative, frequent, and emphasizes the user perspective and experience. In addition to seeking user input, throughout the project lifecycle we will conduct regular evaluation studies involving actual and potential users. Evaluation participants will be recruited through advisory board members' contacts (e.g., colleagues, students in their classes), through our own university's programs, and contacts made through our outreach efforts. We will actively work with our board members and various undergraduate and graduate student organizations to ensure we are also inclusively recruiting under-represented minorities and women into our studies. We will follow our university's IRB guidelines and obtain the necessary approvals for all human-subjects research; generally, these kinds of system evaluation studies are exempt from needing full board approval (due to the low risk posted to participants), but undergo a departmental and university review and approval process. We have completed such a process without incident with our prior Dark Web CRI project.

We will work with our resident HCI expert (Dr. Paul Hu, co-PI) to build in functionality and usability testing throughout the design process. In evaluating system use and experience, the focus will be on user acceptance, satisfaction, task performance, and technology dissemination (Hu et al., 2010). For local users, we will conduct unstructured interviews, focus groups, and lab experiments to assist evaluation. For remote users, we will employ usability testing software and other tools appropriate to the testing task at hand. For usability testing, whether with remote or local users, we will use the Questionnaire for User Interaction Satisfaction (QUIS) (Chin et al., 1988; Slaughter et al., 1995) or other such surveys recommended by our HCI expert, in combination with search and analysis tasks we will ask users to perform. When possible, we will also compare our system to benchmark systems or tools. Quantitative information system and data analytics evaluation metrics such as recall, precision, f-measure, accuracy, task completion, etc. will be adopted. We will perform rigorous statistical tests such as t-test, analysis of variance (ANOVA) test, chi-square test, structured equation modeling, etc. to validate system performances. The effects of using the system will be evaluated at the individual, organization (group), and community levels. The proposed evaluation framework is comprehensive and congruent with the system success model by DeLone and McLean (1992, 2003) and the technology implementation model by Cooper and Zmud (1990). Our system characteristics evaluations are user-centric, thereby ensuring adequate design and implementation of the DIBBs-ISI infrastructure for supporting computational research.

Evaluation results will be funneled to the development team to enhance the system design and implementation. The functionality of the testbed and tools must effectively support information indexing, searches, aggregation, and integration while preserving the necessary privacy and data confidentiality. Accordingly, we will evaluate system functionality along these essential dimensions as well as with respect to the legal or the commonly accepted data access (such as authentication and integrity) and confidentiality requirements.

Ultimately, we will assess the impact of our resource on the research community as a whole, having made it available to the various ISI and CISE research communities, and assess its impacts on knowledge creation and sharing within these communities, using qualitative and quantitative methods and evaluation metrics.

## 6 Outreach, Dissemination, and Education

### 6.1 IEEE ISI, EISIC, PAISI, ISI-KDD and SaTC Meetings

We will leverage our significant contacts within the ISI and SaTC/SFS communities. Several of the project co-PIs and Advisory Board members are key leaders in the community, previously or currently

organizing the IEEE ISI, EISIC, PAISI, and ISI-KDD meetings. Dr. Chen is the founding chair of the ISI series of conferences and has continued to act as steering committee chair of these events. Drs. Yang and Chen will also continue to host the ISI-KDD workshop within the large and popular ACM SIGKDD Conference. In these meetings we will develop special tutorial sessions to introduce the DIBBs-ISI testbed and tools to the community. Additional feedback and suggestions of new contents and services will also be solicited from these meetings. Dr. Chen is also the editor-in-chief of the ACM *Transactions on Management Information Systems* journal and the Springer *Security Informatics* journal. He will actively solicit research that leverages the DIBBs-ISI infrastructure.

Dr. Chen and co-PIs Drs. Yang and Thuraisingham have received current NSF funding within the SaTC and SFS programs for cybersecurity research and education. They will actively promote the DIBBs-ISI infrastructure in the biannual NSF SaTC PI and aspiring PI workshop (about 400 participants). Our proposed testbed and tools, especially hacker forums, IRC, botnet contents, and phishing web sites, will be of great interest to these participants.

### **6.2 Graduate Classes, Minority Recruitment, and SFS and NSA CAE Education Opportunities**

In addition to training a group of graduate students during the DIBBs-ISI development process, all project co-PIs will incorporate elements of our project experience and infrastructure into our respective classes. Drs. Chen and Thuraisingham plan to include the DIBBs-ISI testbed in their classes in Web Computing and Mining (Chen), Security Informatics (Chen), Secure Web Services (Thuraisingham), and Data and Application Security (Thuraisingham). The DIBBs-ISI infrastructure will support hands-on exploration and research on emerging security-related data mining, machine learning, privacy protection, and web analytics topics. Drs. Yang and Abbasi will also include the DIBBs-ISI testbed for their Data Mining and Business Intelligence classes. We will actively recruit minority graduate research assistants from our classes. Dr. Thuraisingham and Larson (both female) will serve as mentors for female students. We will also actively recruit Hispanic-American students due to the larger portion of such students at the University of Arizona in Tucson. We will also develop NSF REU (Research Experience for Undergraduates) proposals for additional undergraduate minority research assistants. Both Drs. Chen and Thuraisingham have mentored many REU students successfully in the past.

Another very promising direction in education will involve incorporating the DIBBs-ISI infrastructure into the high-impact and growing NSF SFS cybersecurity fellowship program and possibly the NSA Center of Academic Excellence (CAE) community in information assurance and security. The NSF SFS program provides funding support through a competitive fellowship for training the next generation of information security professionals. Since 2002, this highly visible NSF program has funded 30+ institutions (e.g., CMU, Purdue, UIUC) that generate an average of 120 SFS students per year. Many of these SFS institutions are also members of the 180+ NSA CAEs, which provide students with NSA-approved cybersecurity curriculums and certificates. The University of Arizona, Drexel University, and UT Dallas are members of the NSA CAE and the NSF SFS programs (Drs. Chen, Yang, and Thuraisingham as project PI or Co-PI). We will promote our DIBBs-ISI infrastructure within the annual NSF SFS PI meeting (next on January 8-10, 2014) and the associated Job Fair for the SFS students (about 400 participants, including government agencies). We believe our infrastructure will serve as an excellent testbed for the many Ph.D. and Master's theses and projects that need to be completed in these SFS programs. We also plan to promote DIBBs-ISI within the NSA CAE listservs and newsletters, as well.

## **7 Project Personnel and Role**

Project funding is requested for selected research collaborators who will be developing, testing, evaluating, stabilizing, and disseminating the testbed portal and infrastructure, as well as for computational and storage facilities. In planning the budget, our priorities were to make as much of the funding available as possible for the actual development and deployment of the DIBBs-ISI infrastructure. We have therefore taken care to minimize the amount of time needed from the PI by including experienced systems staff and students for all development activities, and an experienced project manager whose work will include hands-on as well as supervisory responsibilities. Senior Ph.D. students with significant algorithm and software engineering experience will be assigned to rotate into the project as their specific skills are needed for tool development.

**Dr. Hsinchun Chen** will be project PI and is director of the Artificial Intelligence (AI) Lab. He is an Arizona Regent’s Professor and the Thomas R. Brown Chair Professor at the University of Arizona. The AI Lab currently has 20 staff scientists and research assistants. Chen and his lab have received major funding (\$35M+) from the NSF, NIJ, DARPA, CIA, DHS, and NIH, among others over the past 25 years, and have extensive research experience in the areas of digital government, digital libraries, intelligence and security informatics, health informatics, and knowledge management systems. Dr. Chen was a successful PI in the NSF Digital Library and Digital Government programs. His NSF-funded COPLINK system has been adopted by more than 3,500 agencies and was recently acquired by IBM. Dr. Chen has published more than 250 SCI journal and 150 conference articles and is a leading scholar in security and health informatics. Dr. Chen has significant experience in working with scientists in other disciplines. He will supervise overall progress and lead the infrastructure development effort.

**Dr. Bhavani Thuraisingham** is a Distinguished Professor in the Department of Computer Science at the UT Dallas. Before joining UT Dallas, she had served in the computer industry (MITRE, Honeywell) and NSF (as program director). Her research team has made major breakthroughs in novel class detection for stream mining, adversarial learning, and assured cloud-based information sharing. Over the past seven years, it has generated over \$16 million in research funding and over \$3 million in education funding in cybersecurity. Dr. Thuraisingham will lead in cybersecurity testbed and tool development. **Dr. Chris Yang** is an Associate Professor at the Drexel University and is Associate EIC of the Springer Security Informatics journal. Dr. Chris Yang has served as past IEEE ISI and ISI-KDD program chair and is active in security and health informatics research. In addition to advising DIBBs-ISI development he will take the lead in the ISI and KDD outreach activities. **Dr. Ahmed Abbasi**, Assistant Professor of Information Technology at the University of Virginia, is a former Artificial Intelligence Lab member and an expert in social media analytics and anti-phishing research. He has published in major ACM and IEEE journals and conferences, e.g., IEEE TKDE, ACM TOIS, IEEE JCDL. Dr. Abbasi will work closely with the research team in developing the anti-phishing testbed and tools. **Dr. Paul Hu** of the University of Utah will lead the system evaluation and assessment effort. Dr. Hu has published widely in major information systems journals, especially for healthcare, security and business systems assessment. **Dr. Mark Patton**, Ph.D. in MIS, Director of the Management Information Commons and MicroAge Lab and on the MIS Department’s teaching faculty, will serve as Co-PI with a focus on infrastructure, education, dissemination, and outreach. **Ms. Cathy Larson**, MS in Library Science from UIUC, former librarian, and associate director of the AI Lab, will serve as project manager. She is experienced in portal and digital library development, user evaluation, group facilitation, and system development. She will function as the day-to-day project supervisor and will initiate, coordinate, and collect all advisory board and other user input related to data inclusion and interface design and serve as liaison to the development team.

## 8 Project Planning and Management

### 8.1 Timeline and Milestones

The following timeline and milestones have been planned for this project, and will serve as a guide for ongoing project management. This plan is based on an iterative cycle of development, dissemination, and evaluation that mandates user involvement and feedback throughout each phase of development. The first year, the focus is on development of the initial testbed and on creating the physical infrastructure so that it will support development and concurrent production. In the second year, the focus will be on enhancing the portal including the main user interfaces, and the API and data manager, which will serve as the backend of the portal. This development will continue into the third year. In the third year, the administrative interface will also be developed, in order to support future maintenance and continuity.

**Table 3.** Project Timeline and Tasks

Task Name	Year 1				Year 2				Year 3			
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Project kick-off with advisory board	—————											
Make historical content available	—————				-----				-----			
Initial user feedback on content and tools	—————											

Identify new content and tools	_____		
Develop Information Center	_____		
Incorporate additional contents and tools		_____	
Mngmt. infrastructure & storage server	_____		
User evaluation & feedback	_____		
Testbed API development		_____	
Searcher, Downloader, Analysis Center, DIBBs-ISI portal interfaces		_____	-----
User evaluation & feedback		_____	
Data Manager development/testing		_____	
Administrative interface			_____
Expand Analysis Center to outside users			-----
Stress testing & continuous improvement			_____
Dissemination, updates, closeout			_____

### 8.2 Maintenance and Sustainability

Project funding is not requested for this aspect of the project but is addressed here because it is important that the collection continue to be developed and maintained after the project is over. Maintenance and long-term management will be provided by the staff members of the Artificial Intelligence Lab and the MicroAge Lab after project close. The AI Lab will provide administrative management and systems support using a part-time system administrator. In addition, the MicroAge Lab has dedicated system administration staff to provide additional support as needed. The two groups have close working ties and considerable past NSF project experience in the technologies involved to ensure the quality of service and long-term sustainability of this important resource after project funding has been expended.

### 9 Selected Results from Prior NSF Support

The PI brings over 20 years prior experience to this project. Early work focused on digital libraries and advancements in intelligent information retrieval and naturally progressed to web-based knowledge discovery, information sharing, and security informatics. Many of these projects have relied on the creation of testbeds and other resources that share characteristics with what is proposed here. The proposed project is ambitious, but we believe that the achievements demonstrated in prior work and the significant prototyping work already completed are strong indicators for the high likelihood of success.

- **Hacker Web and AZSecure: Cybersecurity Research and Education (NSF, 2013-2018):** The two projects aim to advance cybersecurity research and education and help develop a new generation of cybersecurity researchers and professionals.
- **Dark Web: Open-Source Terrorism Content Collection and Analysis (NSF, DOD, 2007-2012):** In partnership with selected terrorism study centers and intelligence community agencies, the project aims to develop a systematic, computational approach to collecting and analyzing international extremist-generated Internet contents (web site, forums, chat rooms, etc.).
- **BorderSafe: Border Security Research (NSF, DHS, 2004-2010):** BorderSafe aimed to develop an information sharing and analytics framework for the law enforcement and border security community.
- **BioPortal: National Infectious Disease Infrastructure (NSF, ITIC, 2004-2010):** In partnership with the New York, California, and Arizona public health departments, this project developed a scalable information sharing infrastructure for infectious diseases across jurisdictions and disease types.
- **COPLINK Center: Knowledge Management for Law Enforcement (NSF, NIJ/DOJ, 2000-2006):** A major award of National Institute of Justice and the NSF Digital Government program, this project examined and developed information sharing and crime data mining technologies for public safety. The COPLINK system is in use by more than 3,500 police agencies in twenty states and was acquired by IBM in Summer 2011.

## References

- Abbasi, A., and Chen, H. (2005). "Applying Authorship Analysis to Extremist-Group Web Forum Messages," *IEEE Intelligent Systems*, 20(5), pp. 67-75
- Abbasi, A. and Chen, H. (2007). "Categorization and Analysis of Text in Computer Mediated Communication Archives using Visualization," In the ACM/IEEE Joint Conference on Digital Libraries, Vancouver, Canada, June 18-23, pp. 11-18.
- Abbasi, A. and Chen, H. (2008). "Writeprints: A Stylometric Approach to Identity-Level Identification and Similarity Detection in Cyberspace," *ACM Transactions on Information Systems (TOIS)*, 26:7.
- Abbasi, A. and Chen, H. (2009a) "A Comparison of Fraud Cues and Classification Methods for Fake Escrow Website Detection," *Information Technology and Management*, 10, 2009a, pp. 83-101.
- Abbasi, A. and Chen, H. (2009b). "A Comparison of Tools for Detecting Fake Websites," *IEEE Computer*, 42(10), 2009b, pp. 78-86.
- Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., and Nunamaker Jr., J. F. (2010). "Detecting Fake Websites: The Contribution of Statistical Learning Theory," *MIS Quarterly*, 34(3), 2010, pp. 435 - 461.
- Abbasi, A.; France, S.; Zhang, Z., Chen, H. (2011). "Selecting Attributes for Sentiment Classification Using Feature Relation Networks." *IEEE Transactions on Knowledge and Data Engineering (23:3)*, pp. 447 – 462.
- Abbasi, A., Zahedi, F. M., and Chen, Y. (2012). "Impact of Anti-Phishing Tool Performance on Attack Success Rates," *In the 10th IEEE International Conference on Intelligence and Security Informatics*, Washington D.C., June 11-14, 2012.
- Albertsen, K. (2003). "The Paradigma Web Harvesting Environment," in *Proc. 3rd European Conf. Research and Advanced Tech. for Digital Libraries (ECDL) Workshop on Web Archives*, 2003. <http://bibnum.bnf.fr/ECDL/2003/proceedings.php?f=albertsen>
- Chau, M., Qin, J., Zhou, Y., Tseng, C., and Chen, H. (2005). "SpidersRUs: Automated Development of Vertical Search Engines in Different Domains and Languages," in *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05)*, Denver, Colorado, USA, June 7-11.
- Chau, M., and Xu, J. (2006). "A Framework for Locating and Analyzing Hate Groups in Blogs," in *Proceedings of the Pacific-Asia Conference on Information Systems*, Kuala Lumpur, Malaysia, July 6-9, 2006.
- Chen, H. (2006). *Intelligence and Security Informatics*, Springer-Verlag, 2006.
- Chen, H. (2012). *Dark Web: Exploring and Data Mining the Dark Side of the Web*. Springer, 2012.
- Chen, H., Qin, J., et al. (2004). "Dark Web Portal: Collecting and Analyzing the Presence of Domestic and International Terrorist Groups on the Web." *IEEE Intelligence Transportation Conference*, Washington, D.C., IEEE.
- Chen, H., Larson, C., Elhourani, T., Zimbra, D., Ware, D. (2011a). "The GeoPolitical Web: Assessing Societal Risk in an Uncertain World," *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics, ISI 2011*, Beijing, China, July.
- Chen, H., Denning, D., Roberts, N., Larson, C., Yu, X., and Huang, C.-N. (2011b). "The Dark Web Forum Portal: From Multi-lingual to Video," *Proceedings of the 2011 IEEE International Conference on Intelligence and Security Informatics, ISI 2011*, Beijing, China, July.
- Chin, J.P., Diehl, V.A., & Norman, K.L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. in *CHI '88 Conference Proceedings: Human Factors in Computing Systems*. New York: ACM Press, pp. 213-218.
- Cooper, R.B. and Zmud, R.W. (1990). "Information Technology Implementation Research: A Technology Diffusion Approach." *Management Science* 34:2, pp. 123-139.
- DeLone, W.H. and McLean, E.R. (1992). "Information Systems Success: The Quest for the Dependent Variable." *Information Systems Research* 3(1): 60-95.
- Dhamija, R., Tygar, J. D., and Hearst, M. "Why Phishing Works," *In Proceedings of the ACM International Conference on Computer Human Interaction*, 2006, pp. 581-590.



- Economist, The (2011). "Islam and the Arab Revolutions." March 31, 2011. <http://www.economist.com/node/18486005>.
- Euromoney country risk March 2011: Country Rankings and Acknowledgements (2011). URL: <http://www.euromoney.com/Article/2773235/Country-risk-March-2011-Country-rankings-and-acknowledgements.html>. Accessed 28 Feb. 2011.
- Gartner, *Magic Quadrant for Web Fraud Detection*, April 19, 2011.
- Fayyad, U. M., & Uthurusamy, R. (2002). "Evolving data mining into solutions for insights." *Communications of the ACM*, 45(8), 28-31.
- Fu, T.J., Abbasi, A., Chen, H. (2010). "A Focused Crawler for Dark Web Forums," *Journal of the American Society for Information Science and Technology*, 61:6, pp. 1213-1231.
- Holt, T.J. "Subcultural evolution? Examining the influence of on- and off-line experiences on deviant subcultures," *Deviant Behavior*, 28, 171-198, 2007.
- Holt, T. J., and Kilger, M. "Know Your Enemy: The Social Dynamics of Hacking," *The HoneyNet Project*, 1-17. 2012.
- The HoneyNet Project. *Know your enemy*, 2<sup>nd</sup> Edition. New York: Addison Wesley Professional, 2003.
- Hong, J. "The State of Phishing Attacks," *Communications of the ACM*, 55(1), 2012, pp. 74-81.
- Hu, P., F. M. Hsu, and H. Chen (2010). "Agency Satisfaction with Electronic Record Management Systems: A Large-scale Survey Study." *Journal of the American Society for Information Science and Technology*, 61:12, pp. 2559-2574,.
- Huang, C.-N., Fu, T.J., and Chen, H. (2010). "Text-based Video Content Classification for Online Video-Sharing Sites." *Journal of the American Society for Information Science and Technology*, 61:5, pp. 891-906, 2010.
- Jagatic, T. N., Johnson, N. A., Jakobsson, M., and Menczer, F. "Social Phishing," *Communications of the ACM* (50:10), 2007, pp. 94-100.
- Mielke, Clinton J., Chen, Hsinchun. *Botnet and the Cybercriminal Underground*. IEEE International Conference on Intelligence and Security Informatics, 2008.
- Munson, Z. (2001). "Islamic Mobilization: Social Movement Theory and the Egyptian Muslim Brotherhood." *Sociological Quarterly* 42:4, pp. 487-510.
- Office of Homeland Security. (2002). *National Strategy for Homeland Security*. Washington D.C.: Office of Homeland Security.
- Office of Homeland Security. (2007). *National Strategy for Homeland Security*. Washington, DC: Homeland Security Council.
- Phishtank.com "Phishing Stats," [www.phishtank.com](http://www.phishtank.com), 2013.
- Qin, J., Zhou, Y., and Chau, M. (2004). "Building Domain-Specific Web Collections for Scientific Digital Libraries: A Meta-Search Enhanced Focused Crawling Method," in Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'04), Tucson, Arizona, USA, June 7-11, pp. 135-141.
- Slaughter, L.A., Norman, K.L. & Shneiderman, B. (1995). "Assessing Users' Subjective Satisfaction with the Information System for Youth Services (ISYS)," in Proceedings of Mid Atlantic Human Factors Conference (Blacksburg: March 23-26), pp.22-26.
- Yang, C. and H. Chen (2010), "ISI-KDD: SIGKDD Workshop on Intelligence and Security Informatics 2010," *SIGKDD Explorations*, 12(2), 2010.
- Zhang, Y., X. Yu, Y. Dang, and H. Chen (2010a). "An Integrated Framework for Avatar Data Collection from the Virtual World," *IEEE Intelligent Systems*, 25:5, pp. 17-23.
- Zhang, Y., Zeng, S., Huang, C.-N., Fan, L., Yu, X. Dang, Y., Larson, C., Denning D., Roberts, N., Chen, H. (2010b). "Developing a Dark Web Collection and Infrastructure for Computational and Social Sciences," in Proceedings of the IEEE Intelligence and Security Informatics Conference, pp. 59-64, May 23-26.
- Zhou, Y., Qin, J., Lai, G., Reid, E., and Chen, H. (2006). "Exploring the Dark Side of the Web: Collection and Analysis of U.S. Extremist Online Forums," in Proceedings of the Intelligence and

Security Informatics: IEEE International Conference on Intelligence and Security Informatics (ISI 2006), San Diego, CA, USA, May 23-24, 2006.

Zhou, Y., Reid, E., Qin, J., Chen, H., & Lai, G. (2005). "U.S. Extremist Groups on the Web: Link and Content Analysis." *IEEE Intelligent Systems*, 20:5, 44-51.

Zhu, Z., Lu, G., Chen, Y., Fu, Z. J., Roberts, P., and Han, K. "Botnet Research Survey." *2008 32nd Annual IEEE International Computer Software and Applications Conference*, 967-972, 2008.

**Hsinchun Chen** Regents' Professor and Thomas R. Brown Chair of Management and Technology  
Director, Artificial Intelligence Lab, University of Arizona, Tucson, Arizona

**A. Professional Preparation**

New York University, Information Systems: Ph.D., 1989  
New York University, Information Systems: M.S., 1987  
State University of New York at Buffalo, MIS, Mgmt Science: MBA, 1985

**B. Positions and Honors:**

- Full Professor, University of Arizona, 1998-present.
- IEEE Fellow and AAAS Fellow; IEEE Computer Society Technical Achievement Award, 2006; MISQ Best Paper Award, 2010; IEEE ISI Research Achievement Award, 2011.
- Scientific Counselor/Advisor, National Library of Medicine (USA), Academia Sinica (Taiwan), and National Library of China (China).
- EIC, *ACM Transactions on Management Information Systems* & *Springer Security Informatics*; AE, *Journal of the American Society for Information Science and Technology*, *IEEE Transactions on Systems, Man, and Cybernetics*, and *Decision Support Systems*.
- Steering Committee Chair and Conference Co-Chair for the IEEE Intelligence and Security Informatics Conference (IEEE ISI), 2003-2014. The ISI conference, sponsored by NSF, CIA, DHS, and NIJ, is the premiere meeting for international and homeland security IT research.
- Founding director of Artificial Intelligence Lab and Hoffman E-Commerce Lab. The UA Artificial Intelligence Lab, which houses 20+ researchers, has received more than \$30M in research funding from NSF, DOD, NIH, NLM, DOJ, CIA, and other agencies over the past 20 years.
- Author/ed. of 20 books and more than 250 SCI journal articles and 150 refereed conference papers covering security informatics, biomedical informatics, data/text/web mining, digital library, knowledge management, and Web computing. Selected books include *Dark Web*, 2012; *Sports Data Mining* (2010); *Infectious Disease Informatics* (2010); *Digital Government* (2007); *Intelligence and Security Informatics for International Security* (2006); and *Medical Informatics* (2005); Springer.
- Ranked #8 in publication productivity in Information Systems (CAIS 2005); #1 in Digital Library research (IP&M 2005); h-index of 54, among the top three in MIS (TMIS 2011).
- Dr. Chen's COPLINK system, quoted as a national model for public safety information sharing and analysis, has been adopted in more than 3,500 law enforcement and intelligence agencies. COPLINK research has been featured in the *New York Times*, *Newsweek*, *Washington Post*, and *Boston Globe*, among others. COPLINK was selected as a finalist by the prestigious International Association of Chiefs of Police (IACP)/Motorola 2003 Weaver Seavey Award for Quality in Law Enforcement in 2003. The COPLINK company was acquired by i2 in 2009 and the combined company acquired by IBM in 2011 for \$500M.
- Dr. Chen's Dark Web research, funded by NSF and DOD, has been featured in BBC, Associated Press, National Public Radio, Fox News, *Discover Magazine*, NSF Press, *USA Today*, and the *Washington Post*, among others, as a model advanced computational approach for countering terrorism in cyberspace.
- Received numerous awards in information technology education and industry research including AT&T Foundation Award, SAP Award, the Andersen Consulting Professor of the Year Award, the University of Arizona Technology Innovation Award, and the National Chiao-Tung University Distinguished Alumnus Award.

**C. Selected Publications** (from over 20 books, 250 peer-reviewed journal articles, and 150 refereed conference articles or book chapters): **Most closely related:**

- 1) **H. Chen**. *Dark Web: Exploring and Data Mining the Dark Side of the Web*. Springer, 2012.
- 2) A. Abbasi, S. France, Z. Zhang, and **H. Chen**, "Selecting Attributes for Sentiment Classification Using Feature Relation Networks," *IEEE Trans. on Knowledge and Data Engineering (TKDE)*,

23(3), p. 447-462, 2011.

- 3) A. Abbasi, Z. Zhang, D. Zimbra, **H. Chen**, and J. F. Nunamaker, "Detecting Fake Websites: The Contribution of Statistical Learning Theory," *MIS Quarterly*, 34(3), p. 435-461, 2010. (Winner, MISQ Best paper 2010)
- 4) T. J. Fu, A. Abbasi, and **H. Chen**, "A Focused Crawler for Dark Web Forums," *Journal of the American Society for Information Science and Technology (JASIST)*, 61(6), pp. 1213-1231, 2010.
- 5) D. Hu, S. Kaza, and **H. Chen**, "Identifying Significant Facilitators of Dark Network Evolution," *JASIST*, 60(4), p. 655-665, 2009.

**Other significant works:**

- 6) **H. Chen**, D. Zeng, and P. Yan. *Infectious Disease Informatics: Syndromic Surveillance for Public Health and Bio-Defense*. Springer, 2010.
- 7) **H. Chen** and M. Roco. *Mapping Nanotechnology Innovations and Knowledge: Global and Longitudinal Patent and Literature Analysis*. Springer, 2009.
- 8) **H. Chen**, "AI and Global Science and Technology Assessment," *IEEE Intelligent Systems*, 24(8), p. 68-71, 2009.
- 9) **H. Chen**, E. Reid, J. Sinai, A. Silke, B. Ganor, eds. *Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security*. Springer, 2008.
- 10) A. Abbasi, **H. Chen**, S. Thoms, and T. J. Fu, "Affect Analysis of Web Forums and Blogs using Correlation Ensembles," *IEEE TKDE*, 20(9), p. 1168-1180, 2008.

**D. Synergistic Activities:**

- 1) **HackerWeb and AZSecure** (NSF, \$5.4M). Principal investigator (PI)–PI for these major initiatives; HackerWeb aims to answer the important questions about hacker behaviors, markets, community structure, communication contents, artifacts and cultural differences. AZSecure will support up to 40 students studying advanced cybersecurity analytics and information assurance, for future placement in major government agencies and industries.
- 2) **Dark Web Terrorism Research** (NSF, DTRA, LOC, DHS, \$2.5M). PI. Long-term scientific research program that aims to study modern international terrorism via a computational, data-centric approach; developed the Dark Web Forum Portal which provides searchable access to and social network analysis of terrorist/extremist forums.
- 3) **A National Center of Excellence for Infectious Disease Informatics** (NSF/ITR, 2004-2009, \$2.2M): PI for a collaborative initiative (BioPortal) to explore the development of an integrated and scalable information sharing, monitoring and analysis environment across jurisdictions and for different infectious diseases (e.g., west Nile virus, foot-and-mouth disease).
- 4) **COPLINK Center for Homeland Security Research** (NSF-Digital Government/DHS/CIA, 2000-2007, \$3.1M): PI of this major NSF Digital Government project, which developed information sharing & criminal analysis technologies for the law enforcement & homeland security communities.

**E. Collaborators (Selected):**

Atabakhsh, Homa – Raytheon  
Chen, Ching-chih – Simmons College  
Chen, Su-Shing – The University of Florida  
Demchak, Chris – U.S. Naval War College  
Hu, Paul – University of Utah  
Hubbard, Susan – NIH National Cancer Institute  
Lally, Ann – University of Washington  
Nunamaker, Jay – University of Arizona  
Petersen, Timothy – Tucson Police Dept.  
Ramsey, Marshall – Microsoft  
Roco, Mihail – National Science Foundation

Romano, Nicolas – University of Tulsa  
Scanlon, Pamela – ARJIS (San Diego, Calif.)  
Schatz, Bruce – University of Illinois  
Schroeder, Jenny – Tucson Police Dept.  
Sewell, Robin – University of Arizona  
Violette, Charles – Tucson Police Dept.  
Wilson, Pete – Pima County Sheriff's Dept.  
Wyzga, Wojciech – Knowledge Computing Corp.  
Zhao, Leon – City Univ. of Hong Kong  
**Ph.D. Thesis Advisor:** Vasant Dhar – NYU

## BIOGRAPHICAL SKETCH

NAME <b>Catherine A. Larson</b>	POSITION TITLE Associate Director, Artificial Intelligence Lab Associate Research Scientist		
CONTACT INFORMATION Management Information Systems Dept. The University of Arizona MCCL 430 1130 East Helen Street Tucson, Arizona 85721 USA	520-621-6219 <i>phone</i> 520-626-6499 <i>fax</i> cal@eller.arizona.edu		
EDUCATION: INSTITUTION AND LOCATION University of Illinois at Urbana-Champaign University of Illinois at Urbana-Champaign	DEGREE MSLIS BA	YEAR 1986 1980	FIELD OF STUDY Library & Information Science Spanish; Anthropology (dual major)

### A. POSITIONS AND PROFESSIONAL EXPERIENCE

- Artificial Intelligence Lab, The University of Arizona, Tucson, Arizona, Associate Director and Associate Research Scientist, March 2003 – Present
- The University of Arizona Library, Digital Library Initiatives Group, Data Services Librarian, 1998 – 2003
- The University of Arizona Library, Fine Arts/Humanities Team, Team Leader, 1994 – 1998
- The University of Iowa Libraries, Iowa City, Iowa, Head, Preservation Department, 1990 – 1994
- The University of Iowa Libraries, Media Bibliographer, 1987 – 1990
- The University of Illinois at Urbana-Champaign, Latin American, Collection Development Unit, Visiting Assistant Latin American Bibliographer, 1986

### B. PUBLICATIONS

#### *Selected publications*

1. Victor Benjamin, Wingyan Chung, Ahmed Abbasi, Joshua Chuang, Catherine Larson, Hsinchun Chen. “Evaluating Text Visualization: An Experiment in Authorship Analysis,” in Proc. of the IEEE International Intelligence and Security Informatics Conference, Seattle, Washington, June 4-7, 2013.
2. Hsinchun Chen, Dorothy Denning, Nancy Roberts, Catherine Larson, Ximing Yu, and Chun-Neng Huang. “The Dark Web Forum Portal: From Multi-lingual to Video,” in Proc. of the IEEE International Intelligence and Security Informatics Conference, Beijing, China, July 10-12, 2011.
3. Yulei Zhang, Shuo Zeng, Chun-neng Huang, Li Fan, Ximing Yu, Yan Dang, Catherine Larson, Dorothy Denning, Nancy Roberts, and Hsinchun Chen, “Developing a Dark Web Collection and Infrastructure for Computational and Social Sciences,” in Proc. of the IEEE International Conference on Intelligence and Security Informatics, ISI 2010, Vancouver, Canada, May 2010.
4. Xuan Liu, P. Zhang, Xin Li, Hsinchun Chen, Yan Dang, Y. Lin, Catherine Larson, Mihail C. Roco, and X. Wang “Trends for Nanotechnology Development in China, Russia, and India,” *Journal of Nanoparticle Research*, 11:8, p. 1845-1866, November, 2009.
5. Hsinchun Chen, Lawrence Brandt, Valerie Gregg, Roland Traummüller, Sharon Dawes, Eduard Hovy, Ann Macintosh and Catherine Larson, eds. (2008). *Digital Government: Advanced Research and Case Studies*. NY: Springer.

6. Xin Li, Daning Hu, Yan Dang, Hsinchun Chen, Mihail C. Roco, Catherine Larson, and Joyce Chan, "Nano Mapper: An Internet Knowledge Mapping System for Nanotechnology Development," *Journal of Nanoparticle Research*, Volume 10, 2008.
7. Paul Hu, Daniel Zeng, Catherine Larson, Wei Chang, Chunju Tseng, Jian Ma, and Hsinchun Chen (2007). "System for Infectious Disease Information Sharing and Analysis: Design and Evaluation," *IEEE Transactions on Information Technology in Biomedicine* 11(4), July 2007.
8. Paul Jen-Hwa Hu, Daniel Zeng, Hsinchun Chen, Catherine Larson, Wei Chang, Chunju Tseng (2005). "Evaluating an Infectious Disease Information Sharing and Analysis System." Proceedings, IEEE Intelligence & Security Informatics, in Lecture Notes in Computer Science 3495.
9. Betsy Kruger and Catherine Larson, eds. (2006). *On Account of Sex: An Annotated Bibliography on the Status of Women in Librarianship, 1998-2002*. Lanham MD: Scarecrow Press.
10. Homa Atabakhsh, Catherine Larson, Tim Petersen, Chuck Violette, and Hsinchun Chen (2004). "Information Sharing and Collaboration Policies with Government Agencies." Proceedings, 2nd Symposium in Intelligence & Security Informatics, in Lecture Notes in Computer Science 3073.

### C. SELECTED GRANTS AND AWARDS

1. National Science Foundation (NSF), "Cybersecurity Scholarship-for-Service at The University of Arizona." Co-writer and SFS Program Coordinator for the University of Arizona. PI: H. Chen. \$4.2M.
2. NSF, Community Resource Development (CRD), "Developing a Dark Web Collection and Infrastructure for Computational and Social Sciences," 2007 – 2010. Co-PI, and principle writer and coordinator for writing, project planning, and project implementation. PI: H. Chen. \$500K.
3. Institute for Museum and Library Services, to The University of Arizona Library for creation of the Arizona Electronic Atlas, a GIS service. Writing and project team member, 2001-03. PI: Chris Kollen. \$129K.
4. American Library Association, Association of College & Research Libraries, Instruction Section. Member of team (the University of Arizona Library's Education Project Team) that received the "Innovation in Instruction" award for creating an innovative tutorial for information literacy, 1999.
5. With Jim Fromm and Steve Llewellyn, "Scenario Planning for Facilities Management and Planning. At: Society for College and University Planning, Mid-Continent Regional Conference, Tucson AZ, October 1999. (Awarded "Best of Regional Conference" and selected for presentation at National Conference, Denver, Colorado, July 2000).
6. National Endowment for the Humanities, to the Iowa Cooperative Preservation Consortium for state-wide preservation planning grant. Member of planning and writing team and assistant project director, 1994- 95. \$37,000.

### D. COLLABORATORS (Selected)

Hsinchun Chen, U. of Arizona  
 Michael Ascher, LLNL  
 Susan Brown, U. Of Arizona  
 Wei Chang, U. of Arizona  
 Yan (Mandy) Dang, U. Northern Ariz.  
 Denning, Dorothy, Naval Postgrad. School  
 Millicent Eidson, NY Dept. of Health  
 Ivan Gotham, NY Dept. of Health  
 Daning Hu, U. of Arizona

Paul Hu, U. of Utah  
 Betsy Kruger, UIUC  
 Xin Li, City U. of Hong Kong  
 Jian Ma, Cataphora Inc.  
 Nancy Roberts, Naval Postgrad. School  
 Mihail Roco, NSF  
 Chunju Tseng, U. of Arizona  
 Daniel Zeng, U. of Arizona  
 Yulei (Gavin) Zhang, U. Northern Ariz.

## BIOGRAPHICAL SKETCH

NAME <b>Mark W. Patton</b>	POSITION TITLE Lecturer, Associate Director, Hoffman E-Commerce Lab		
INSTITUTION AND LOCATION	DEGREE	YEAR(s)	FIELD OF STUDY
University of Arizona	Ph.D.	2009	MIS
University of Michigan	MBA	1999	Management
Montana College of Mineral Science & Technology	B.S	1990	Computer Science & Mining Engineering

### a. Professional and Academic Positions and Appointments

- 2009-present    Lecturer: Management Information Systems
- Information Security, Risk Management, and Disaster Recovery (MIS 516)
  - The Strategic Management of Information Systems (MIS 113)
  - Computers and The Internetworked Society (MIS 111)
- 2005-2009      Adjunct Lecturer: Management Information Systems
- Project Management (BAD 520) Executive MBA course
  - Computers and The Internetworked Society (MIS 111)
  - Data and Network Communications (MIS 307)
  - Systems Analysis and Design (MIS 363)
  - Guest lectures in Entrepreneurship on E-Commerce Business Planning and Operations Management on Operations Simulation
- 2002-2009      Manager, Hoffman E-Commerce Lab, Management Information Systems Department, University of Arizona, Tucson, Arizona
- Engage in hosting resources, conducting IS research, supporting faculty research activities, and providing advanced IT resources to faculty and graduate students.
  - Coordinate collaborative educational, research, and outreach activities.
  - Co-host business entrepreneurship seminars for local business entrepreneurs with SCORE (Senior Core of Retired Executives).
  - Manage IS projects, professional staff, and student GTA's.
  - Engage in development and advancement activities with key companies to identify collaborative research projects and elicit support for ongoing initiatives.
  - Engage in research on educational technologies and applications.
- 2002, 2003      Sponsorship Co-Chair, Intelligence and Security Informatics Symposium, Tucson, Arizona
- 1999-2001      Senior E-Business Associate, DiamondCluster International Inc., Chicago, Illinois
- Coordinated community and personalization aspects of financial services portal including origination, development, testing, and vendor management.
  - Developed the personalization data-structure and programmed the personalization rules for financial services portal.
  - Developed use cases and testing sequences for financial services and B2C service industry portals.
  - Developed and consulted on on-line privacy policies in the financial sector.
  - Identified, analyzed, and reported on business strategies and strategic e-business opportunities in manufacturing and finance.
  - Performed business development for internet start-up ventures, developing industry penetration strategies. Successfully recruited key industry representative for client's board of directors.

- Summer 1998 Internal Consultant, DaimlerChrysler, Sterling Heights, Michigan
- Applied lean manufacturing techniques to save \$6MM on new vehicle launch.
  - Modeled process flows and benchmarked competitive production systems to identify optimized location for \$3MM in water test facilities.
- 1994-1997 Senior Engineer and Production Supervisor, FMC Wyoming Corporation, Green River, Wyoming
- Administered systems and networks in heterogeneous UNIX and Windows environment.
  - Designed and implemented WAN connecting servers and workstations at remote locations with a central database server.
  - Coordinated \$45MM annual and \$6-\$10MM capital budgeting processes at regional production center (mine).
  - Developed and coordinated operational plans, streamlined operations, and laid the groundwork for future mine expansions.
  - Managed capital improvement projects of \$3.2MM, coordinated development and implementation across departmental boundaries.
  - Supervised and motivated underground production and maintenance crews.
  - Led United Way campaign to record employee participation rate, increasing year-over-year from 44% to 64% at unionized 1,150 employee location
- 1990-1994 Engineer and Senior Engineer, FMC Wyoming Corporation, Kemmerer, Wyoming
- Programmed a new production and maintenance tracking and reporting system.
  - Developed operational plans and coordinated the operations of \$15MM surface coal mine, increased recoverable reserves and extended the life of the mine by 3 years.
  - Devised and implemented capital projects of \$2.2MM
  - Interviewed and trained new engineers in engineering, budgeting, forecasting, and operations coordination.

## **b. Publications**

Patton, M. W.; Nunamaker, J. F.; Neumann; W. T.; Burgoon, J. K. (2010) "Developing a Prototype Kiosk for Delivering Deception Detection Decision Support" in Hawaii International Conference on System Science (HICSS'10) Hawaii 2010.

Patton, M. W. (2009) "Automated Assessment of Truth and Deception in a Rapid Assessment Environment using Kiosk Based Avatars" Dissertation

Patton, M. W.; Patton, W. E., and Dalton, L. S. (1994) StockTracker Simulation, Glencoe-McGraw Hill.

## **c. Synergistic Activities**

- PI – CITeR Grant: Rapid Assessment Kiosk based Interviews
- Ongoing research into Deception Detection using Embodied Avatar's in a Kiosk Environment



## AHMED ABBASI

Associate Professor of Information Technology  
University of Virginia  
Email: [abbasi@comm.virginia.edu](mailto:abbasi@comm.virginia.edu)  
Phone: (434) 924-7031

---

### EDUCATION

---

**Ph.D. in Management Information Systems** with minor in Cognitive Science, 2008  
Eller College of Management, University of Arizona, Tucson

**Masters in Business Administration** (MBA-IT Concentration), 2004  
Pamplin College of Business, Virginia Tech, Blacksburg

**Bachelor of Science** (Business Information Technology), 2002  
Pamplin College of Business, Virginia Tech, Blacksburg

---

### WORK EXPERIENCE

---

**Associate Professor, Information Technology:**

2014-present, McIntire School of Commerce, University of Virginia

- Conduct research and teach courses related to web/text mining and security.
- Director, Center for Analytics

**Assistant Professor, Information Technology:**

2011-present, McIntire School of Commerce, University of Virginia

- Conducted research and taught courses related to web/text mining and security.

**Assistant Professor, Management Information Systems:**

2008-2011, Sheldon B. Lubar School of Business, University of Wisconsin, Milwaukee (UWM)

- Conducted research related to text mining, taught undergraduate and graduate level programming courses

**Research Associate and Project Lead:**

2004-2008, Artificial Intelligence Lab, University of Arizona, Tucson

- Managed team of researchers involved in creating large digital libraries and online text archives.

**Research Assistant:**

2003, Department of Business Information Technology, Virginia Tech, Blacksburg

- Conducted research related to the use of Neural Networks for data mining applications.

---

### PUBLICATIONS CLOSELY RELATED TO PROJECT

---

1. [Abbasi, A.](#) and Chen, H. "A Comparison of Tools for Detecting Fake Websites," *IEEE Computer*, 42(10), 2009, pp. 78-86.
2. [Abbasi, A.](#), Zahedi, F. M., and Kaza, S. "Detecting Fake Medical Web Sites using Recursive Trust Labeling," *ACM Transactions on Information Systems*, 30(4), 2012, no. 22.
3. [Abbasi, A.](#), Zahedi, F. M., and Chen, Y. "Impact of Anti-Phishing Tool Performance on Attack Success Rates," In the 10<sup>th</sup> *IEEE International Conference on Intelligence and Security Informatics*, Washington D.C., June 11-14, 2012.
4. [Abbasi, A.](#), Zhang, Z., Zimbra, D., Chen, H., and Nunamaker Jr., J. F. "Detecting Fake Websites: The Contribution of Statistical Learning Theory," *MIS Quarterly*, 34(3), 2010, pp. 435-461 ([MISQ Best Paper Award for 2010](#), [AIS 5 Best Publications for 2010](#)).
5. Fu, T., [Abbasi, A.](#), and Chen, H. "A Focused Crawler for Dark Web Forums," *Journal of the American Society for Information Science and Technology*, 61(6), 2010, pp. 1213-1231.

---

## OTHER SIGNIFICANT PUBLICATIONS

---

1. Abbasi, A., and Chen, H. "Identification and Comparison of Extremist-Group Web Forum Messages using Authorship Analysis," *IEEE Intelligent Systems*, 20(5), 2005, pp. 67-75.
2. Abbasi, A., and Chen, H. "Writeprints: A Stylometric Approach to Identity-Level Identification and Similarity Detection in Cyberspace," *ACM Transactions on Information Systems*, 26(2), 2008, no. 7.
3. Kopeke, J., Kaza, S., and Abbasi, A. "Identification of Malicious and Fake Websites: Using Information from the Transport and Internet Layers of the Network Stack," In the *10<sup>th</sup> IEEE International Conference on Intelligence and Security Informatics*, Washington D.C., June 11-14, 2012.
4. Abbasi, A. and Chen, H. "A Comparison of Fraud Cues and Classification Methods for Fake Escrow Website Detection," *Information Technology and Management*, 10(2), 2009, pp. 83-101.
5. Fu, T., Abbasi, A., Zeng, D., and Chen, H. "Sentimental Spidering: Leveraging Opinion Information in Focused Crawlers," *ACM Transactions on Information Systems*, 30(4), 2012, no. 24.

---

## SYNERGISTIC ACTIVITIES

---

1. Associate Editor, *ACM Transactions on MIS* (2013 – present); *IEEE Intelligent Systems* (2014 – present)
2. Senior Member, IEEE.
3. Director, Center for Analytics, McIntire School of Commerce, UVA (2013 – present).
4. Coordinator, Enterprise IT Module, MS MIT Program, McIntire School of Commerce, UVA (2011 – present).
5. Member, UWM's Research Cyber-Infrastructure Working Group (2008-2009), UWM's Medical Informatics Program (2008-2011).
6. Dissertation committee member:  
Nitin Walia (UWM, 2010), Craig Claybaugh (UWM, 2010), Yan Chen (UWM, 2012).
7. Program committee member:  
Annual Meeting of the Association for Computational Linguistics (ACL 2010, 2011), Conference of the North American Chapter of the ACL (NAACL 2013), Conference of the European Chapter of the ACL (EACL 2012), International Joint Conference on Natural Language Processing (IJCNLP 2011), Pacific Asia Workshop on Intelligence and Security Informatics (PAISI 2011), International Workshop on Search and Mining User-generated Contents (SMUC 2010), Workshop on Information Technologies and Systems (WITS 2009-2013).
8. Reviewed manuscripts for the following journals:  
*IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Systems, Man, and Cybernetics*, *IEEE Computer*, *IEEE Intelligent Systems*, *ACM Transactions on Information Systems*, *ACM Transactions on Knowledge and Data Discovery*, *Journal of the American Society for Information Science and Technology*, *Decision Support Systems*, *Management Science*, *MIS Quarterly*, *Information Systems Research*, *Journal of MIS*, *Information Processing and Management*, *Journal of Computer-Mediated Communication*, *Language Resources and Evaluation*.

---

## COLLABORATORS AND OTHER AFFILIATIONS

---

### 1. Recent Collaborators

Donald Adjeroh, West Virginia University  
Conan Albrecht, Brigham Young University  
Victor Benjamin, University of Arizona  
Hsinchun Chen, University of Arizona  
Yan Chen, University of Auburn-Montgomery  
Wingyan Chung, UNC-Fayetteville  
Tianjun Fu, Google Inc.  
James V. Hansen, Brigham Young University  
Ammar Hassan, University of Virginia

Siddharth Kaza, Towson University  
Cathy Larson, University of Arizona  
Jay F. Nunamaker Jr., University of Arizona  
Anthony Vance, Brigham Young University  
Mariam Zahedi, University of Wisconsin-Milwaukee  
Daniel Zeng, University of Arizona  
Zhu Zhang, University of Arizona  
David Zimbra, Santa Clara University

## Paul Jen-Hwa Hu

David Eccles Professor of Information Systems

Department of Operations and Information Systems, David Eccles School of Business, University of Utah

Telephone: (801) 587-7785; Fax: (801) 581-7731; Email: paul.hu@business.utah.edu

Personal Web page: [https://faculty.utah.edu/u0239890-Paul\\_Hu/biography/index.html](https://faculty.utah.edu/u0239890-Paul_Hu/biography/index.html)

### **Education**

August 1993 – May 1998	Ph.D. in Management Information Systems, University of Arizona
August 1989 – August 1992	M.S. in Management Information Systems, University of Arizona
September 1985 – June 1987	Bachelor of Science in Chemical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C.

### **Professional Experiences and Appointments**

July 2011 – Present	David Eccles Professor of Information Systems, University of Utah
July 2009 – Present	Professor, University of Utah
July 2009 – June 2011	David Eccles Scholar, Professor, University of Utah
July 2003 – June 2009	Associate Professor (tenured), David Eccles Faculty Fellow, U. of Utah
August 2000 – June 2003	Assistant Professor, University of Utah
January 1998 – July 2000	Assistant Professor, University of South Florida

**Research Interests:** Large-scale System Evaluations and Impact Assessments; Information Technology in Health Care; E-Commerce; Digital Government; Information Technology Implementation Management; Human-Computer Interactions; Business Intelligence; Knowledge Management.

### **Selected Journal Publications Closely Related to the Proposed Project**

1. Y. Dang, Y. Zhang, P. J. Hu, S. Brown, Y. Ku, J. Wang, and H. Chen, “An Integrated Framework for Analyzing Multilingual Content in Web 2.0 Social Media,” *Decision Support Systems* (forthcoming).
2. Y. Dang, Y. Zhang, H. Chen, S. Brown, P. J. Hu, and J. Nunamaker, “Theory-Informed Design and Evaluation of An Advanced Search and Knowledge Mapping System in Nanotechnology,” *Journal of Management Information Systems*, Vol.28, No.3, 2012, pp.101-129.
3. Y. Dang, Y. Zhang, P. J. Hu, S. Brown, and H. Chen, “Knowledge Mapping for Rapidly Evolving Domains: A Design Science Approach,” *Decision Support Systems*, Vol.50, No.2, 2011, pp.415-427.
4. Y. Dang, Y. Zhang, H. Chen, P. J. Hu, S. Brown, and C. Larson, “Arizona Literature Mapper: An Integrated Approach to Monitor and Analyze Global Bioterrorism Research Literature,” *Journal of the American Society for Information Science and Technology*, Vol.60, No.7, 2009, pp.1466-1485.
5. P. J. Hu, D. Zeng, H. Chen, C. Larson, W. Chang, and C. Tseng, “A System for Infectious Disease Information Sharing and Analysis: Design, Implementation and Evaluation,” *IEEE Transactions on Information Technology in Biomedicine*, Vol.11, No.4, 2007, pp.483-492.

### **Selected Other Significant Journal Publications**

1. F. Xiao, P. J. Hu, L. Li, and W. Tsai, “Predicting Adoption Probabilities in Social Networks,” *Information Systems Research*, Vol.24, No.1, 2013, pp.128-145.
2. X. Fang, P. J. Hu, M. Chau, H. Hu, Z. Yang, and O. Liu Sheng, “A Data-Driven Approach to Measure Web Site Navigability,” *Journal of Management Information Systems*, Vol. 29, No.2, 2012, pp.173-212.
3. Y. Chen, S. Brown, P. J. Hu, C. King, and H. Chen, “Managing Emerging Infectious Diseases with Information Systems: Reconceptualizing Outbreak Management through the Lens of Loose Coupling,” *Information Systems Research*, Vol.22, 2011, pp.447-468.

4. P. J. Hu, F. Hsu, H. Hu, and H. Chen, "Agency Satisfaction with Electronic Record Management Systems: A Large-Scale Survey Study," *Journal of the American Society for Information Science and Technology*, Vol.61, No.12, 2010, pp.2559-2574.
5. C. Wei, P. J. Hu, and Y. Lee, "Preserving User Preferences in Automated Document-Category Management: An Evolution-Based Approach," *Journal of Management Information Systems*, Vol.25, No.4, 2009, pp.109-143.

### **Recent Research Grants**

1. X. Fang, and P. J. Hu, *Seed Grants*, University of Utah, \$29,000, 2013-2014.
2. Paul J. Hu, National Center of Excellence for Infectious Disease Informatics, *NSF IRA* (IIS 0428241), \$75,000 (Y413911, University of Arizona, total funding \$1,250,016), 2005-2009.
3. O. Liu Sheng and P. J. Hu, SERG, *NSF Digital Government* (Award ID 0410409), \$25,000, 2004-2005.

### **Synergistic Activities**

1. Designing and conducting large-scale evaluations of Web 2.0 portals and information systems in public health, security informatics, and knowledge management domains, and sharing important experiences with the respective research communities through conferences and online content.
2. Working with health care professionals, public health specialists, and law enforcement experts to refine systems designs and evaluations that involve real-world users, thereby producing effective information systems that benefit the current practices.
3. Developing and evaluating novel computational methods to analyze message/opinion disseminations and product adoptions by individuals in an online social network.
4. Disseminating practice system design considerations and evaluation experiences in graduate-level information systems classes.

### **Collaborators**

1. Hsinchun Chen, University of Arizona
2. Chih-Ping Wei, National Taiwan University
3. Xiao Fang, University of Utah
4. Patrick Chau, University of Hong Kong
5. Han-fen Hu, University of Nevada, Las Vegas
6. Bryan Bonner, University of Utah
7. Sue Brown, University of Arizona
8. Olivia Sheng, University of Utah

**Dissertation Advisor:** Olivia Sheng (formerly at University of Arizona, currently at University of Utah)

### **Doctoral Student Advisees**

1. Han-fen Hu; received Ph.D. degree from University of Utah (2012); currently Assistant Professor at University of Nevada, Las Vegas
2. Jongtae Yu, Ph.D. student at University of Utah
3. Dissertation committee member of Lionel Li, University of Utah Ph.D. graduate (2013); Iljoo Kim, University of Utah Ph.D. graduate (2011).

## BIOGRAPHICAL SKETCH

**Name:** Bhavani Thuraisingham  
**Business Address:** Erik Jonsson School of Engineering and Computer Science  
The University of Texas at Dallas, Richardson, TX 75080  
Telephone: (972) 883 - 4738 (voice), 883-2399 (FAX)  
Email: Bhavani.thuraisingham@utdallas.edu  
<http://www.utdallas.edu/~bxt043000/> [www.dr-bhavani.org](http://www.dr-bhavani.org)

### EDUCATION

University of Ceylon (1st Class)	Mathematics and Physics	B.S. (1975)
The University of Bristol, UK	Mathematics	M.S. (1977)
University of Wales, UK	Computer Science	Ph.D. (1979)
University of Bristol, UK	Engineering	D.Eng (2011)

### PROFESSIONAL EDUCATION (1980-PRESENT)

MS in Computer Science, U of MN, GPA 4.0, 1983; Java Programming Certificate, Learning Tree International, 2000; Certificate in Terrorism Studies, St. Andrews University, Scotland, 2010; ISC2 Certified Information Systems Security Professional, 2010.

### EXPERIENCE

9/10-Present: Louis Beecherl, Jr. I Distinguished Professor, UT Dallas  
10/04 – 9/10: Professor of Computer Science and Director of Cyber Security Research Center  
10/01-9/04 Program Director, the National Science Foundation  
Information and Data Management ('02), Data and Applications Security ('03), Cyber Trust ('04)  
1/89-6/05 The MITRE Corporation  
Information Technology Consultant (10/04 – 6/05); NSF/IPA (10/01-9/04); Chief  
Scientist/Engineer (5/99-9/01), Department Head, (6/95 – 5/99), Principal Scientist/Engineer  
(3/95-6/95, Lead Scientist/Engineer (1/89 – 3/95)  
6/10-present Founder, Knowledge and Security Analytics, Software Development  
6/05 – present President, Bhavani Security Consulting (IT and security consulting)  
8/99-5/01 Adjunct professor of Computer Science, Boston University Metropolitan College  
1/86 – 1/89 Principal Research Scientist, Honeywell Inc  
12/83 - 1/86 Senior Analyst, Control Data Corporation  
9/84 – 12/88 Adjunct professor of Computer Science and Member of Graduate Faculty, Univ. of MN  
9/81 – 6/82 Visiting Professor of Mathematics, University of Minnesota  
8/80 – 5/81 Visiting Professor of Computer Science, New Mexico Institute of Technology

### FIVE RELATED PUBLICATIONS (out of 100+ journals, 200+ conferences, 12 books 3 patents)

1. Tyrone Cadenhead, Murat Kantarcioglu, Bhavani M. Thuraisingham: Scalable and Efficient Reasoning for Enforcing Role-Based Access Control. DBSec 2010
2. Barbara Carminati, Elena Ferrari, Raymond Heatherly, Murat Kantarcioglu, Bhavani M. Thuraisingham: A semantic web based framework for social network access control. SACMAT 2009.
3. Bhavani M. Thuraisingham: Privacy constraint processing in a privacy-enhanced database management system. Data Knowl. Eng. 55(2) (2005)
4. Bhavani M. Thuraisingham: On the Complexity of the Privacy Problem in Databases. Data Mining: Foundations and Practice 2008
5. Li Liu, Murat Kantarcioglu, Bhavani M. Thuraisingham: The applicability of the perturbation based privacy preserving data mining for real-world data. Data Knowl. Eng. 65(1) (2008)

### FIVE OTHER PUBLICATIONS

1. B. Thuraisingham, Database and Applications Security, CRC Press, 2005
2. Elisa Bertino, Barbara Carminati, Elena Ferrari, Bhavani M. Thuraisingham, Amar Gupta: Selective and Authentic Third-Party Distribution of XML Documents. IEEE Trans. Knowl. Data Eng. 16 (2004)
3. K. Hamlen, B. Thuraisingham, M. Kantarcioglu, L. Khan, Security Issues for Cloud Computing, Journal of Information Security and Privacy Journal, Volume 4, December 2010.
4. Manghui Tu, Peng Li, I-Ling Yen, Bhavani M. Thuraisingham, Latifur Khan: Secure Data Objects Replication in Data Grid. IEEE Trans. Dependable Sec. Comput. 7(1) (2010)

5. Anna Cinzia Squicciarini, Elisa Bertino, Elena Ferrari, Federica Paci, Bhavani M. Thuraisingham: PP-trust-X: A system for privacy preserving trust negotiations. ACM Trans. Inf. Syst. Secur. 10(3): (2007)

### **SYNERGISTIC ACTIVITIES**

1. IFIP WG11.3 Working Group on Data and Applications Security (1990 – present) and Member, IEEE Computer Society Committees (1997-Present)
2. Editorial Boards-Sample: IEEE Transactions on Dependable and Secure Computing, ACM Transactions on Information and Systems Security, Journal of Information Security, Journal of Semantic Web; Journal of Computer Security, IEEE Transactions on Knowledge and Data Eng.
3. Editor in Chief: Computer Standards and Interface Journal, 2005-2009;
4. Advisory Boards: Department of Computer Science Purdue University, Journal of Privacy Technologies; member of panels for Air Force Scientific Advisory Board and the National Academy of Sciences
5. Served as Program Chair for over 15 conferences and keynote presenter at over 80 conferences; Editor of over 10 books and several conference/workshop proceedings

### **HONORS AND AWARDS**

1. Fellow of IEEE (2003), AAAS (2003) and BCS (British Computer Society, 2005)
2. IEEE Distinguished Lecturer (2002-2005)
3. ACM Distinguished Scientist (2010)
4. ACM SIGSAC Outstanding Contribution Award for seminal research contributions and leadership in data and applications security for over 25 years, October 2010
5. Research Leadership Award in Intelligence and Security Informatics, jointly presented by IEEE SMC and IEEE ITSC, May 2010
6. IEEE Computer Society's 1997 Technical Achievement Award, "For outstanding and innovative contributions to Secure Distributed Data Management" August 1997.
7. Featured in Silicon India Magazine as one of leading seven technology innovators of South Asian Origin, May 2002
8. Woman of Color Technology Research Leadership award, Career Communications Inc., Fall 2001.

**PROFESSIONAL ACCOMPLISHMENTS** Author of 100+ journal papers, 3 patents granted in database inference control and 1 pending on stream mining, 12 books (10 authored, 2 coauthored) including Database and Applications Security; Presentations on Data Mining, National Security and Privacy at the White House Office of Science and Technology Policy and the United Nations

### **LIST OF COLLABORATORS & CO-EDITORS WITHIN THE PAST 48 MONTHS**

E. Bertino (Purdue), B. Bhargava (Purdue), S. Das (NSF/UTA), S. Jajodia (GMU), B. Carminati (U of Como), C. Clifton (Purdue), E. Ferrari (U of Como), T. Finin (UMBC), J. Han (UIUC), S. Iyengar (LSU), A. Joshi (UMBC), K. Kim (UC Irvine), S. Sahni (UF), R. Sandhu (GMU), S. Shekhar (UMN), M. Winslett (UIUC), S. Yau (ASU), Faculty from UTDallas, Members from MITRE Corporation

### **ADVISEES AND POSTDOCTORAL SCHOLARS SPONSORED (PAST FIVE YEARS)**

**Students prior to joining UTD:** Co-supervised 1 PhD student at University of Milan (B. Carminati, 2000-2004); Co-supervised three PhD students at the University of Minnesota (K. Hwang. T. Keefe, S. Chen, 1985-1988); Co-supervised 2 MS students at the University of Rhode Island

**Current PhD students at UTD:**, 1., Ontology matching; 2. Jyothsna Rachapalli, Policy Management, 3. Vaibhav Khadilkar, Managing Large RDF Graphs

**PhD Students Graduated:** Li Liu (Privacy), R. Layfield (Information Sharing), Zhong Wang (Security), Tyrone Cadenhead (Inference Control), Wei-She (Secure SOA), Neda Alipanah (Semantic Web) M. Farhan Husain (Secure Cloud), Jeff Partyka (Geospatial Data Management) Parveen Pallabi (

**Post-docs at UTD:** Tyrone Cadenhead (Inference Control)

**MS Students at UTD:** Twelve MS students graduated in Data Applications and Security

### **NAMES OF GRADUATE AND POSTDOCTORAL ADVISORS**

Dr. John Cleave, University of Bristol, UK; Dr. Roger Hindley, University of Wales, UK

## **Christopher C. Yang, Associate Professor**

**College of Computing and Informatics**

**Drexel University**

**Email: [chris.yang@drexel.edu](mailto:chris.yang@drexel.edu)**

### **Education**

**Ph.D., Electrical and Computer Engineering**, University of Arizona, Tucson, May, 1997.

**M.S., Electrical and Computer Engineering**, University of Arizona, Tucson, May, 1992.

**B.S., Electrical and Computer Engineering**, University of Arizona, Tucson, August, 1990.

### **Work Experience**

**College of Computing and Informatics, Drexel University**

*June 2008 – present: Associate Professor*

**Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong**

*January 2001 – May 2008: Associate Professor*

*January 2000 – December 2000: Assistant Professor*

*January 2000 – May 2008: Director of Digital Library Laboratory*

**Department of Computer Science and Information Systems, University of Hong Kong**

*September 1997 – December 1999: Assistant Professor*

*September 1997 – December 1999: Director of Digital Library Research Laboratory*

*January 1998 – December 1999: Associate Director of Authorized Academic Java<sup>SM</sup> Campus<sup>SM</sup> (AAJC)*

**Selected Publications:** (88 journal papers and 147 conference papers, a complete list can be found at <http://cci.drexel.edu/faculty/cyang/pubs.html>)

(5 related)

1. X. Tang and C. C. Yang, "Detecting Social Media Hidden Communities using Dynamic Stochastic Blockmodel with Temporal Dirichlet Process," *ACM Transactions on Intelligent Systems and Technology*, accepted for publication.
2. Z. Hai, K. Chang, J. Ki, and C. C. Yang, "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance," *IEEE Transactions on Knowledge and Data Engineering*, vol.26, no.3, March 2014, pp.623-634.
3. X. Tang and C. C. Yang, "Ranking User Influence in Healthcare Social Media," *ACM Transactions in Intelligent Systems and Technology*, vol.3, no.4, 2012.
4. C. C. Yang and T. Ng, "Analyzing and Visualizing Web Opinion Development and Social Interactions with Density Based Clustering," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol.41, no.6, 2011, pp.1144-1155.
5. C. C. Yang, X. Shi, and C. Wei, "Discovering Event Evolution Graphs from News Corpora," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol.39, no.4, 2009, pp.850-863.

(5 additional)

6. C. C. Yang and X. Tang, "Estimating User Influence in the MedHelp Social Network," *IEEE Intelligent Systems*, vol.27, no.5, 2012.
7. X. Tang and C. C. Yang, "TUT: A Statistical Model for Detecting Trends, Topics and User Interests in Social Media," *Proceedings of the 21st ACM Conference on Information and Knowledge Management, Maui, October 29 - November 2, 2012.*

8. H. Yang and C. C. Yang, "Harnessing Social Media for Drug-Drug Interactions Detection," *Proceedings of IEEE International Conference on Healthcare Informatics*, Philadelphia, PA, September 8 - 11, 2013.
9. X. Tang, M. Zhang, and C. C. Yang, "Leveraging User Interest to Improve Thread Recommendation in Online Forum," *Proceedings of International Conference on Social Intelligence and Technology*, State College, PA, May 8 – 9, 2013.
10. X. Tang, M. Zhang, and C. C. Yang, "User Interest and Topic Detection for Personalized Recommendation," *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence*, Macau, December 4-7, 2012.

### **Selected Professional Services:**

1. General co-chair, International Conference on Healthcare Informatics 2013, Philadelphia, Pennsylvania (ICHI 2013)
2. General co-chair, International Conference on Social Intelligence and Technology 2013, College Station, Pennsylvania (SOCIETY 2013)
3. Program Committee Chair, ACM SIGHIT International Health Informatics Symposium, Miami, FL, USA, 2012.
4. Program Co-chair, IEEE International Conference on Intelligence and Security Informatics, Vancouver, BC, Canada, 2010.
5. Workshop co-chair, ACM CIKM International Workshop on Smart Health and Wellbeing, Maui, Hawaii (SHB 2012)
6. Workshop Co-chair, ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD), D.C., July, 2010.
7. Track Co-chair, Web 2.0, and Social Media Analytics Track, International Conference on Information Systems (ICIS), Phoenix, AZ, USA, 2009.
8. Program Vice-chair, ACM Eighteenth Conference on Information and Knowledge Management (CIKM), Hong Kong, China, 2009.
9. Guest editor, ACM Transactions on Management Information Systems, Special Issue on Informatics for Smart Health and Wellbeing
10. Guest editor, ACM Transactions on Intelligent Systems and Technology, Special Issues on Intelligent Systems for Health Informatics
11. Guest editor of *IEEE Transactions on Systems, Man, & Cybernetics, Part A*, Special Issue on Social Media Analytics: Understanding the Pulse of the Society.
12. Associate Editor-in-Chief, Security Informatics, Springer
13. Co-Editor, Electronic Commerce Research & Applications, Elsevier

**Collaborators and Co-Editors:** Sophia Ananiadou (University of Manchester, UK), Carlo Combi (Universita' degli Studi di Verona, Italy), Uwe Glaesser (Simon Fraser U., Canada), Wai Lam (Chinese U. of Hong Kong, HK), Ee-Peng Lim (Singapore Management U., Singapore), Jiming Liu (Hong Kong Baptist University), Zhiyong Lu (National Library of Medicine), Yin-Leng Theng (Nanyang Technological University, Singapore), Bhavani Thuraisingham (U. of Texas, Dallas, USA), Jennifer Unger (University of Southern California), Ke Wang (Simon Fraser U., Canada), Chih-Ping Wei (National Tsing Hua U., Taiwan), Flaura Winston (U. of Pennsylvania, USA), John Yen (Pennsylvania State University), Daniel Zeng (U. of Arizona, USA), Qing Zeng (University of Utah)

**Post-Graduate Scholar Sponsor:** Tobun D. Ng, V. Naikan, Shanfeng Zhu.

**Ph.D. and Master Thesis Advisor: Current:** Mi Zhang, Haodong Yang, Ling Jiang, Haemin Kim.  
**Graduated:** Katherine Chuang 12, Xuning Tang 12, Anthony Ng 09, Ying-Chau Wong 08, Jianfeng Lin 07, Nan Liu 06, Xiaodong Shi 06, Tung-Yin Cheng 05, Kar-Wing Li, 04, Patricia Chan 04, Rachel Wong 04, Fu-Lee Wang 03, Milo Yip 02, Mandy Chan 01, Winter Chan 01, Kay Hong 01, Alan Chung 00, Johnny Luk 00, Stanley Yung 00.



# SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION <b>University of Arizona</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Hsinchun Chen</b>				AWARD NO.	Proposed	Granted	
				A. SENIOR PERSONNEL: PI/PI, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)			
				CAL	ACAD	SUMR	
1. <b>Hsinchun Chen - PI</b>				0.00	0.00	0.75	<b>22,506</b>
2. <b>Catherine A Larson - Co-PI</b>				1.00	0.00	0.00	<b>6,009</b>
3. <b>Mark W Patton - Co-PI</b>				0.00	0.00	1.00	<b>8,800</b>
4.							
5.							
6. ( <b>0</b> ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00	<b>0</b>
7. ( <b>3</b> ) TOTAL SENIOR PERSONNEL (1 - 6)				1.00	0.00	1.75	<b>37,315</b>
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00	<b>0</b>
2. ( <b>0</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				7.20	0.00	0.00	<b>51,000</b>
3. ( <b>3</b> ) GRADUATE STUDENTS							<b>55,000</b>
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS							<b>0</b>
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)							<b>0</b>
6. ( <b>0</b> ) OTHER							<b>0</b>
TOTAL SALARIES AND WAGES (A + B)							<b>143,315</b>
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)							<b>68,400</b>
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)							<b>211,715</b>
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT							<b>0</b>
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							<b>2,400</b>
2. FOREIGN							<b>2,000</b>
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____ <b>0</b>							
2. TRAVEL _____ <b>0</b>							
3. SUBSISTENCE _____ <b>0</b>							
4. OTHER _____ <b>0</b>							
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> ) TOTAL PARTICIPANT COSTS							<b>0</b>
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							<b>4,500</b>
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION							<b>0</b>
3. CONSULTANT SERVICES							<b>0</b>
4. COMPUTER SERVICES							<b>0</b>
5. SUBAWARDS							<b>171,746</b>
6. OTHER							<b>0</b>
TOTAL OTHER DIRECT COSTS							<b>176,246</b>
H. TOTAL DIRECT COSTS (A THROUGH G)							<b>392,361</b>
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 51.5000, Base: 287921)</b>							
TOTAL INDIRECT COSTS (F&A)							<b>148,279</b>
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)							<b>540,640</b>
K. RESIDUAL FUNDS							<b>0</b>
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)							<b>540,640</b>
M. COST SHARING PROPOSED LEVEL \$ <b>0</b> AGREED LEVEL IF DIFFERENT \$							
PI/PD NAME <b>Hsinchun Chen</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
				Date Checked	Date Of Rate Sheet	Initials - ORG	

# SUMMARY PROPOSAL BUDGET

YEAR **2**

ORGANIZATION <b>University of Arizona</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Hsinchun Chen</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
		CAL	ACAD	SUMR			
1.	<b>Hsinchun Chen - PI</b>	0.00	0.00	0.75	<b>22,506</b>		
2.	<b>Catherine A Larson - Co-PI</b>	1.00	0.00	0.00	<b>6,009</b>		
3.	<b>Mark W Patton - Co-PI</b>	0.00	0.00	1.00	<b>8,800</b>		
4.							
5.							
6.	( <b>0</b> ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	<b>0</b>		
7.	( <b>3</b> ) TOTAL SENIOR PERSONNEL (1 - 6)	1.00	0.00	1.75	<b>37,315</b>		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1.	( <b>0</b> ) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	<b>0</b>		
2.	( <b>0</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	7.20	0.00	0.00	<b>51,000</b>		
3.	( <b>3</b> ) GRADUATE STUDENTS				<b>55,000</b>		
4.	( <b>0</b> ) UNDERGRADUATE STUDENTS				<b>0</b>		
5.	( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				<b>0</b>		
6.	( <b>0</b> ) OTHER				<b>0</b>		
TOTAL SALARIES AND WAGES (A + B)					<b>143,315</b>		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					<b>68,400</b>		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					<b>211,715</b>		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT					<b>0</b>		
E. TRAVEL							
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)					<b>4,400</b>		
2. FOREIGN					<b>2,000</b>		
F. PARTICIPANT SUPPORT COSTS							
1.	STIPENDS \$ _____	<b>0</b>					
2.	TRAVEL _____	<b>0</b>					
3.	SUBSISTENCE _____	<b>0</b>					
4.	OTHER _____	<b>0</b>					
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> )				TOTAL PARTICIPANT COSTS	<b>0</b>		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					<b>3,500</b>		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					<b>0</b>		
3. CONSULTANT SERVICES					<b>0</b>		
4. COMPUTER SERVICES					<b>0</b>		
5. SUBAWARDS					<b>160,555</b>		
6. OTHER					<b>0</b>		
TOTAL OTHER DIRECT COSTS					<b>164,055</b>		
H. TOTAL DIRECT COSTS (A THROUGH G)					<b>382,170</b>		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 51.5000, Base: 199311)</b>							
TOTAL INDIRECT COSTS (F&A)					<b>102,645</b>		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					<b>484,815</b>		
K. RESIDUAL FUNDS					<b>0</b>		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					<b>484,815</b>		
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME <b>Hsinchun Chen</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

# SUMMARY PROPOSAL BUDGET

YEAR 3

ORGANIZATION <b>University of Arizona</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Hsinchun Chen</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
		CAL	ACAD	SUMR			
1.	<b>Hsinchun Chen - PI</b>	0.00	0.00	0.75	<b>22,506</b>		
2.	<b>Catherine A Larson - Co-PI</b>	1.00	0.00	0.00	<b>6,009</b>		
3.	<b>Mark W Patton - Co-PI</b>	0.00	0.00	1.00	<b>8,800</b>		
4.							
5.							
6.	( 0 ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	<b>0</b>		
7.	( 3 ) TOTAL SENIOR PERSONNEL (1 - 6)	1.00	0.00	1.75	<b>37,315</b>		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1.	( 0 ) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	<b>0</b>		
2.	( 0 ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	7.20	0.00	0.00	<b>51,000</b>		
3.	( 3 ) GRADUATE STUDENTS				<b>55,000</b>		
4.	( 0 ) UNDERGRADUATE STUDENTS				<b>0</b>		
5.	( 0 ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				<b>0</b>		
6.	( 0 ) OTHER				<b>0</b>		
TOTAL SALARIES AND WAGES (A + B)					<b>143,315</b>		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					<b>68,400</b>		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					<b>211,715</b>		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
	<b>Production server</b>		\$	<b>5,500</b>			
TOTAL EQUIPMENT					<b>5,500</b>		
E. TRAVEL							
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)					<b>4,400</b>		
2. FOREIGN					<b>2,000</b>		
F. PARTICIPANT SUPPORT COSTS							
1.	STIPENDS \$ _____			<b>0</b>			
2.	TRAVEL _____			<b>0</b>			
3.	SUBSISTENCE _____			<b>0</b>			
4.	OTHER _____			<b>0</b>			
TOTAL NUMBER OF PARTICIPANTS ( 0 )				TOTAL PARTICIPANT COSTS	<b>0</b>		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					<b>3,000</b>		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					<b>0</b>		
3. CONSULTANT SERVICES					<b>0</b>		
4. COMPUTER SERVICES					<b>0</b>		
5. SUBAWARDS					<b>147,749</b>		
6. OTHER					<b>0</b>		
TOTAL OTHER DIRECT COSTS					<b>150,749</b>		
H. TOTAL DIRECT COSTS (A THROUGH G)					<b>374,364</b>		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
<b>MTDC (Rate: 51.5000, Base: 193616)</b>							
TOTAL INDIRECT COSTS (F&A)					<b>99,712</b>		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					<b>474,076</b>		
K. RESIDUAL FUNDS					<b>0</b>		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					<b>474,076</b>		
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME <b>Hsinchun Chen</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

# SUMMARY PROPOSAL BUDGET

Cumulative

ORGANIZATION <b>University of Arizona</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Hsinchun Chen</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PI, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
		CAL	ACAD	SUMR			
1.	<b>Hsinchun Chen - PI</b>	0.00	0.00	2.25	<b>67,518</b>		
2.	<b>Catherine A Larson - Co-PI</b>	3.00	0.00	0.00	<b>18,027</b>		
3.	<b>Mark W Patton - Co-PI</b>	0.00	0.00	3.00	<b>26,400</b>		
4.							
5.							
6.	( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	<b>0</b>		
7.	( <b>3</b> ) TOTAL SENIOR PERSONNEL (1 - 6)	3.00	0.00	5.25	<b>111,945</b>		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1.	( <b>0</b> ) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	<b>0</b>		
2.	( <b>0</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	21.60	0.00	0.00	<b>153,000</b>		
3.	( <b>9</b> ) GRADUATE STUDENTS				<b>165,000</b>		
4.	( <b>0</b> ) UNDERGRADUATE STUDENTS				<b>0</b>		
5.	( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				<b>0</b>		
6.	( <b>0</b> ) OTHER				<b>0</b>		
TOTAL SALARIES AND WAGES (A + B)					<b>429,945</b>		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					<b>205,200</b>		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					<b>635,145</b>		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
				<b>\$ 5,500</b>			
TOTAL EQUIPMENT					<b>5,500</b>		
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)					<b>11,200</b>		
2. FOREIGN					<b>6,000</b>		
F. PARTICIPANT SUPPORT COSTS							
1.	STIPENDS \$ _____				<b>0</b>		
2.	TRAVEL _____				<b>0</b>		
3.	SUBSISTENCE _____				<b>0</b>		
4.	OTHER _____				<b>0</b>		
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> ) TOTAL PARTICIPANT COSTS					<b>0</b>		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					<b>11,000</b>		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					<b>0</b>		
3. CONSULTANT SERVICES					<b>0</b>		
4. COMPUTER SERVICES					<b>0</b>		
5. SUBAWARDS					<b>480,050</b>		
6. OTHER					<b>0</b>		
TOTAL OTHER DIRECT COSTS					<b>491,050</b>		
H. TOTAL DIRECT COSTS (A THROUGH G)					<b>1,148,895</b>		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
TOTAL INDIRECT COSTS (F&A)					<b>350,636</b>		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					<b>1,499,531</b>		
K. RESIDUAL FUNDS					<b>0</b>		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					<b>1,499,531</b>		
M. COST SHARING PROPOSED LEVEL \$ <b>0</b> AGREED LEVEL IF DIFFERENT \$							
PI/PI NAME <b>Hsinchun Chen</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

C \*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

## **Budget Justification**

### **I. Personnel:**

#### **PI:**

Hsinchun Chen, Ph.D.: PI, 0.75 SU month (6.25% time). Dr. Chen is Regent's Professor and endowed Thomas R. Brown Chair in Management and Technology at the University of Arizona. As the Director of the Artificial Intelligence Lab, he will serve as project PI, establish research directions, ensure that project milestones are met, and project and university resources used most effectively. As a leading expert in the intelligence and security informatics field, he will also have a key role in ensuring community involvement and input throughout the project.

#### **Co-PIs:**

Mark Patton, Ph.D.: Co-PI, 1.00 SU month (8.33% time). Patton is Director of the Management Information Commons and MicroAge Lab and on the MIS Department's teaching faculty, with a specialty in information assurance and cybersecurity. Patton's Co-PI role will include working closely with the PI on research directions, advising on the hardware infrastructure and enhancements needed over the life of the project as storage needs increase and as user demand grows, and integrating the infrastructure with the Microage and AI Lab server, storage, and back-up facilities for permanent hosting. He will incorporate the tools and use of the data into the classes he teaches as well as make them available to other faculty.

Ms. Catherine Larson: Co-PI, 1.00 cal. mo. (8.33% time). As Co-PI, Larson will function as day to day project manager and supervisor. She will also have a lead role in designing and coordinating systems evaluation research including user studies, and will serve as liaison to the University Institutional Review Board (IRB). She will work closely with the PI to facilitate communication with the collaborating partners and on following up with community outreach and involvement. Ms. Larson has substantive management and coordination experience in AI Lab research projects, digital library projects, user studies, and data protection. She will also collect and interpret advisory board feedback and serve as liaison to the development team. Like Patton, she will also help with dissemination and outreach efforts.

#### **Other Project Staff:**

Systems engineer, 7.2 cal mo. (60% time). The systems engineer will have primary responsibility for system and software development, database development and administration, and system integration. The systems engineer will work closely with Dr. Patton on setting up the most efficient and effective system architecture, and with the graduate students as they develop and/or enhance the tools and archive through programming and algorithm development. His overall responsibilities include development, testing, bug fixes, and operationalizing and moving the resources and tools to production mode for use.

Graduate Students: Ph.D. and masters students assigned to this work will have the expertise, skills, and experience needed to ensure project success including data collection and web spidering; algorithm development for text, data mining, and web mining; social network analysis; visualization development. They will work closely with Dr. Patton, the systems engineer, and with the contracted Co-PIs and their staff members. More junior students will be assigned to the project to assist the senior Ph.D. students, especially with tasks such as data collection and parsing, and system architecture and development, and to be mentored by them. We expect that several part-time Ph.D. students will work on this project, rotating in as their expertise is needed. In addition, part-time

masters students with coding, programming, database administration, and/or user interface design and development skills will support the PhD students' work. A total of 1.10 FTE (110% time) of UA student time will be employed, with the time apportioned among the individual students as they are brought into the project.

## **II. Fringe Benefit Rates:**

Faculty and Appointed Staff: Chen, Patton, Larson: 30.0%

Classified Staff: Systems engineer, 48.0%

Graduate Students: Ph.D. and masters students, 59.5% (including tuition remission at 50.0%)

Actual rates in place during the time of the award will be charged.

## **III. Capitalized equipment:**

Funding is requested in the first year for a development server; estimates are approximate, based on currently available equipment. The development server will be used for all initial system development efforts followed by system integration of the tools and data. It will also support early prototyping and testing with users and others. Funding is requested in the third year for a production server, at which point we will be ready to begin hardening some of the system components for regular use. The production server will be used to serve all tools no longer undergoing active regular development. The development server will continue to be used throughout the project for all new tools and data.

The specific brands and models to be purchased will be specified at the time of award in order to take advantage of the most current technologies and prices.

## **IV. Travel:**

Travel funds are requested to support PI and other project staff travel for community involvement and outreach, workshops and presentations for purposes of demonstrating the archive and research infrastructure, and for testing with users we might otherwise not have access to.

## **V. Other Direct Costs:**

### *Subcontracts:*

The following subcontracts will be awarded:

- University of Texas at Dallas, Bhavani Thuraisingham, Ph.D., PI. – Dr. Thuraisingham's team will lead in cybersecurity testbed and tool development.
- Drexel University, Chris Yang, Ph.D., PI. – In addition to advising on development, Dr. Yang will also lead in the ISI and KDD outreach activities.
- University of Virginia, Ahmed Abbasi, Ph.D., PI. – Dr. Abbasi will contribute a significant amount of data and work closely with the PhD and masters students on algorithm development.
- University of Utah, Paul Hu, Ph.D., PI. – Dr. Hu will design user and system evaluations and work with project manager Larson on evaluation implementation.

More detailed information about project roles can be found in the Project Description. Detailed budget information including committed project time, salaries, fringe benefits, and indirect cost rates, and other information can be found in the Budgets and Budget Justifications of the respective subcontracts.

*Materials & supplies:* Funding is requested for storage media, reference manuals, software licenses, and miscellaneous small equipment (such as memory upgrades, hard drives, and the like) which are specific to this project. Considerable data storage may be needed.

## **VI. Indirect Costs:**

Indirect Costs are calculated in accordance with the university's federally negotiated indirect cost rate. The indirect cost rate is charged against the Modified Total Direct Cost (MTDC), which includes all direct costs except the tuition remission portion of fringe benefits and capitalized equipment, and only the first \$25,000 per subcontract. The indirect cost rate for the University of Arizona is 51.5% for all years.

Actual rates in place during the time of the award will be charged.

# SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION <b>Drexel University</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Chris Yang</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
				CAL	ACAD	SUMR	
1. <b>Chris Yang - PI</b>				0.00	0.00	1.00	<b>13,666</b>
2.							
3.							
4.							
5.							
6. ( 0 ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00	<b>0</b>
7. ( 1 ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	1.00	<b>13,666</b>
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( 0 ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00	<b>0</b>
2. ( 0 ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00	<b>0</b>
3. ( 1 ) GRADUATE STUDENTS							<b>12,500</b>
4. ( 0 ) UNDERGRADUATE STUDENTS							<b>0</b>
5. ( 0 ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)							<b>0</b>
6. ( 0 ) OTHER							<b>0</b>
TOTAL SALARIES AND WAGES (A + B)							<b>26,166</b>
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)							<b>4,441</b>
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)							<b>30,607</b>
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT							<b>0</b>
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							<b>1,000</b>
2. FOREIGN							<b>0</b>
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____				<b>0</b>			
2. TRAVEL _____				<b>0</b>			
3. SUBSISTENCE _____				<b>0</b>			
4. OTHER _____				<b>0</b>			
TOTAL NUMBER OF PARTICIPANTS ( 0 ) TOTAL PARTICIPANT COSTS							<b>0</b>
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							<b>0</b>
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION							<b>0</b>
3. CONSULTANT SERVICES							<b>0</b>
4. COMPUTER SERVICES							<b>0</b>
5. SUBAWARDS							<b>0</b>
6. OTHER							<b>0</b>
TOTAL OTHER DIRECT COSTS							<b>0</b>
H. TOTAL DIRECT COSTS (A THROUGH G)							<b>31,607</b>
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 54.5000, Base: 31607)</b>							
TOTAL INDIRECT COSTS (F&A)							<b>17,226</b>
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)							<b>48,833</b>
K. RESIDUAL FUNDS							<b>0</b>
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)							<b>48,833</b>
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME <b>Chris Yang</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked		Date Of Rate Sheet		Initials - ORG	



# SUMMARY PROPOSAL BUDGET

YEAR 2

ORGANIZATION <b>Drexel University</b>				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Chris Yang</b>				AWARD NO.	Proposed	Granted
					NSF Funded Person-months	
A. SENIOR PERSONNEL: PI/PI, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. <b>Chris Yang - PI</b>				0.00	0.00	1.00
2.						
3.						
4.						
5.						
6. ( 0 ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. ( 1 ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	1.00
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. ( 0 ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00
2. ( 0 ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00
3. ( 1 ) GRADUATE STUDENTS						12,500
4. ( 0 ) UNDERGRADUATE STUDENTS						0
5. ( 0 ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						0
6. ( 0 ) OTHER						0
TOTAL SALARIES AND WAGES (A + B)						26,576
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						4,575
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						31,151
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						0
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						1,200
2. FOREIGN						0
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____				0		
2. TRAVEL _____				0		
3. SUBSISTENCE _____				0		
4. OTHER _____				0		
TOTAL NUMBER OF PARTICIPANTS ( 0 ) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						0
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						0
3. CONSULTANT SERVICES						0
4. COMPUTER SERVICES						0
5. SUBAWARDS						0
6. OTHER						0
TOTAL OTHER DIRECT COSTS						0
H. TOTAL DIRECT COSTS (A THROUGH G)						32,351
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 54.5000, Base: 32351)</b>						
TOTAL INDIRECT COSTS (F&A)						17,631
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						49,982
K. RESIDUAL FUNDS						0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						49,982
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$		
PI/PI NAME <b>Chris Yang</b>				FOR NSF USE ONLY		
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION		
		Date Checked	Date Of Rate Sheet	Initials - ORG		

# SUMMARY PROPOSAL BUDGET

YEAR 3

ORGANIZATION <b>Drexel University</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Chris Yang</b>				AWARD NO.	Proposed	Granted	
				A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)			
				CAL	ACAD	SUMR	
1. <b>Chris Yang - PI</b>				0.00	0.00	1.00	<b>14,498</b>
2.							
3.							
4.							
5.							
6. ( 0 ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00	<b>0</b>
7. ( 1 ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	1.00	<b>14,498</b>
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( 0 ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00	<b>0</b>
2. ( 0 ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00	<b>0</b>
3. ( 1 ) GRADUATE STUDENTS							<b>12,500</b>
4. ( 0 ) UNDERGRADUATE STUDENTS							<b>0</b>
5. ( 0 ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)							<b>0</b>
6. ( 0 ) OTHER							<b>0</b>
TOTAL SALARIES AND WAGES (A + B)							<b>26,998</b>
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)							<b>4,712</b>
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)							<b>31,710</b>
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT							<b>0</b>
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							<b>1,400</b>
2. FOREIGN							<b>0</b>
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____ <b>0</b>							
2. TRAVEL _____ <b>0</b>							
3. SUBSISTENCE _____ <b>0</b>							
4. OTHER _____ <b>0</b>							
TOTAL NUMBER OF PARTICIPANTS ( 0 ) TOTAL PARTICIPANT COSTS							<b>0</b>
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							<b>0</b>
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION							<b>0</b>
3. CONSULTANT SERVICES							<b>0</b>
4. COMPUTER SERVICES							<b>0</b>
5. SUBAWARDS							<b>0</b>
6. OTHER							<b>0</b>
TOTAL OTHER DIRECT COSTS							<b>0</b>
H. TOTAL DIRECT COSTS (A THROUGH G)							<b>33,110</b>
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 54.5000, Base: 33110)</b>							
TOTAL INDIRECT COSTS (F&A)							<b>18,045</b>
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)							<b>51,155</b>
K. RESIDUAL FUNDS							<b>0</b>
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)							<b>51,155</b>
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME <b>Chris Yang</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked		Date Of Rate Sheet		Initials - ORG	

# SUMMARY PROPOSAL BUDGET Cumulative

ORGANIZATION <b>Drexel University</b>				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Chris Yang</b>				AWARD NO.	Proposed	Granted
				NSF Funded Person-months		
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. <b>Chris Yang - PI</b>				0.00	0.00	3.00
2.						
3.						
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. ( <b>1</b> ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	3.00
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00
2. ( <b>0</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00
3. ( <b>3</b> ) GRADUATE STUDENTS						37,500
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS						0
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						0
6. ( <b>0</b> ) OTHER						0
TOTAL SALARIES AND WAGES (A + B)						79,740
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						13,728
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						93,468
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						0
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						3,600
2. FOREIGN						0
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____ <b>0</b>						
2. TRAVEL _____ <b>0</b>						
3. SUBSISTENCE _____ <b>0</b>						
4. OTHER _____ <b>0</b>						
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> ) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						0
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						0
3. CONSULTANT SERVICES						0
4. COMPUTER SERVICES						0
5. SUBAWARDS						0
6. OTHER						0
TOTAL OTHER DIRECT COSTS						0
H. TOTAL DIRECT COSTS (A THROUGH G)						97,068
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)						
TOTAL INDIRECT COSTS (F&A)						52,902
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						149,970
K. RESIDUAL FUNDS						0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						149,970
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$		
PI/PD NAME <b>Chris Yang</b>				FOR NSF USE ONLY		
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION		
		Date Checked	Date Of Rate Sheet	Initials - ORG		

C \*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

**Christopher C. Yang, Associate Professor**  
**College of Computing and Informatics**  
**Drexel University**  
**Email: [chris.yang@drexel.edu](mailto:chris.yang@drexel.edu)**

## **Budget Justification**

### Project Personnel:

- PI for Drexel University: Christopher C. Yang, Ph.D. Yang's team will work closely with project PI Chen to advise on testbed development and to help build, create, and contribute tools. He will also actively promote the DIBBs-ISI infrastructure at the ISI and PAISI workshops. He will also serve as a liaison to the KDD community to ensure awareness of the availability of the data. He will also include the testbed in his classes on business intelligence, data mining, and other security- and computing-related classes to which it is relevant. Students in his classes will also be solicited to serve as potential participants for the user and system evaluation studies. Yang will spend approximately 1 month per academic year on the project.
- Graduate Assistants for Drexel will, under Yang's direction and guidance, help build and create tools for the infrastructure. They will also help the Arizona team with ongoing data collection efforts, in order to keep the test bed as up to date as possible, and will serve as early testers for the system. Two quarter-time assistants will be hired specifically for this project.

### Fringe Benefits:

- Fringe benefits are calculated at the current university allowable rates of 32.50% for PI Yang. Rates actually in effect at the time of the project will be charged against the project.

### Travel:

- Travel funding for the KDD and ISI/PAISI conferences has been budgeted allowing Yang to actively serve as liaison and presenter.

### Indirect Costs (IDC):

- Drexel's IDC rate is presently 54.50%; the rate in effect during the project is what will actually be charged. The IDC is calculated by multiplying the modified total direct costs (MTDC) by the IDC rate.

# SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION <b>University of Texas at Dallas</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Bhavani Thuraisingham</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR				
1. <b>Bhavani Thuraisingham - PI</b>	0.00	0.00	0.25	<b>4,861</b>			
2.							
3.							
4.							
5.							
6. ( <b>0</b> ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	<b>0</b>			
7. ( <b>1</b> ) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.25	<b>4,861</b>			
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	<b>0</b>			
2. ( <b>2</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.50	<b>6,805</b>			
3. ( <b>1</b> ) GRADUATE STUDENTS				<b>15,750</b>			
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS				<b>0</b>			
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				<b>0</b>			
6. ( <b>0</b> ) OTHER				<b>0</b>			
TOTAL SALARIES AND WAGES (A + B)				<b>27,416</b>			
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				<b>5,279</b>			
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				<b>32,695</b>			
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT				<b>0</b>			
E. TRAVEL				<b>2,000</b>			
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							
2. FOREIGN				<b>0</b>			
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS	\$		<b>0</b>				
2. TRAVEL			<b>0</b>				
3. SUBSISTENCE			<b>0</b>				
4. OTHER			<b>0</b>				
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> )				TOTAL PARTICIPANT COSTS		<b>0</b>	
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES				<b>0</b>			
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				<b>0</b>			
3. CONSULTANT SERVICES				<b>0</b>			
4. COMPUTER SERVICES				<b>0</b>			
5. SUBAWARDS				<b>0</b>			
6. OTHER				<b>0</b>			
TOTAL OTHER DIRECT COSTS				<b>0</b>			
H. TOTAL DIRECT COSTS (A THROUGH G)				<b>34,695</b>			
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
<b>MTDC (Rate: 53.0000, Base: 34696)</b>							
TOTAL INDIRECT COSTS (F&A)				<b>18,389</b>			
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				<b>53,084</b>			
K. RESIDUAL FUNDS				<b>0</b>			
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				<b>53,084</b>			
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME <b>Bhavani Thuraisingham</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

# SUMMARY PROPOSAL BUDGET

YEAR 2

ORGANIZATION <b>University of Texas at Dallas</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Bhavani Thuraisingham</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR				
1. <b>Bhavani Thuraisingham - PI</b>	0.00	0.00	0.25		<b>5,007</b>		
2.							
3.							
4.							
5.							
6. ( <b>0</b> ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00		<b>0</b>		
7. ( <b>1</b> ) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.25		<b>5,007</b>		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS	0.00	0.00	0.00		<b>0</b>		
2. ( <b>0</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00		<b>0</b>		
3. ( <b>1</b> ) GRADUATE STUDENTS					<b>15,750</b>		
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS					<b>0</b>		
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)					<b>0</b>		
6. ( <b>0</b> ) OTHER					<b>0</b>		
TOTAL SALARIES AND WAGES (A + B)					<b>20,757</b>		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					<b>3,614</b>		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					<b>24,371</b>		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT					<b>0</b>		
E. TRAVEL					<b>2,065</b>		
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							
2. FOREIGN					<b>0</b>		
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____					<b>0</b>		
2. TRAVEL _____					<b>0</b>		
3. SUBSISTENCE _____					<b>0</b>		
4. OTHER _____					<b>0</b>		
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> )				TOTAL PARTICIPANT COSTS	<b>0</b>		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					<b>0</b>		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					<b>0</b>		
3. CONSULTANT SERVICES					<b>0</b>		
4. COMPUTER SERVICES					<b>0</b>		
5. SUBAWARDS					<b>0</b>		
6. OTHER					<b>0</b>		
TOTAL OTHER DIRECT COSTS					<b>0</b>		
H. TOTAL DIRECT COSTS (A THROUGH G)					<b>26,436</b>		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
<b>MTDC (Rate: 53.0000, Base: 26436)</b>							
TOTAL INDIRECT COSTS (F&A)					<b>14,011</b>		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					<b>40,447</b>		
K. RESIDUAL FUNDS					<b>0</b>		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					<b>40,447</b>		
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME <b>Bhavani Thuraisingham</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

# SUMMARY PROPOSAL BUDGET

YEAR 3

ORGANIZATION <b>University of Texas at Dallas</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Bhavani Thuraisingham</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR				
1. <b>Bhavani Thuraisingham - PI</b>	0.00	0.00	0.13		<b>2,579</b>		
2.							
3.							
4.							
5.							
6. ( <b>0</b> ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00		<b>0</b>		
7. ( <b>1</b> ) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.13		<b>2,579</b>		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS	0.00	0.00	0.00		<b>0</b>		
2. ( <b>0</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00		<b>0</b>		
3. ( <b>1</b> ) GRADUATE STUDENTS					<b>10,500</b>		
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS					<b>0</b>		
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)					<b>0</b>		
6. ( <b>0</b> ) OTHER					<b>0</b>		
TOTAL SALARIES AND WAGES (A + B)					<b>13,079</b>		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					<b>2,221</b>		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					<b>15,300</b>		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT					<b>0</b>		
E. TRAVEL					<b>2,000</b>		
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							
2. FOREIGN					<b>0</b>		
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS	\$		<b>0</b>				
2. TRAVEL			<b>0</b>				
3. SUBSISTENCE			<b>0</b>				
4. OTHER			<b>0</b>				
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> )				TOTAL PARTICIPANT COSTS	<b>0</b>		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					<b>0</b>		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					<b>0</b>		
3. CONSULTANT SERVICES					<b>0</b>		
4. COMPUTER SERVICES					<b>0</b>		
5. SUBAWARDS					<b>0</b>		
6. OTHER					<b>0</b>		
TOTAL OTHER DIRECT COSTS					<b>0</b>		
H. TOTAL DIRECT COSTS (A THROUGH G)					<b>17,300</b>		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 53.0000, Base: 17300)</b>							
TOTAL INDIRECT COSTS (F&A)					<b>9,169</b>		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					<b>26,469</b>		
K. RESIDUAL FUNDS					<b>0</b>		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					<b>26,469</b>		
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME <b>Bhavani Thuraisingham</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

# SUMMARY PROPOSAL BUDGET Cumulative

ORGANIZATION <b>University of Texas at Dallas</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Bhavani Thuraisingham</b>				AWARD NO.	Proposed	Granted	
				A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)			
				CAL	ACAD	SUMR	
1. <b>Bhavani Thuraisingham - PI</b>				0.00	0.00	0.63	<b>12,447</b>
2.							
3.							
4.							
5.							
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00	<b>0</b>
7. ( <b>1</b> ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	0.63	<b>12,447</b>
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00	<b>0</b>
2. ( <b>2</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.50	<b>6,805</b>
3. ( <b>3</b> ) GRADUATE STUDENTS							<b>42,000</b>
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS							<b>0</b>
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)							<b>0</b>
6. ( <b>0</b> ) OTHER							<b>0</b>
TOTAL SALARIES AND WAGES (A + B)							<b>61,252</b>
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)							<b>11,114</b>
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)							<b>72,366</b>
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT							<b>0</b>
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							<b>6,065</b>
2. FOREIGN							<b>0</b>
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____ <b>0</b>							
2. TRAVEL _____ <b>0</b>							
3. SUBSISTENCE _____ <b>0</b>							
4. OTHER _____ <b>0</b>							
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> ) TOTAL PARTICIPANT COSTS							<b>0</b>
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							<b>0</b>
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION							<b>0</b>
3. CONSULTANT SERVICES							<b>0</b>
4. COMPUTER SERVICES							<b>0</b>
5. SUBAWARDS							<b>0</b>
6. OTHER							<b>0</b>
TOTAL OTHER DIRECT COSTS							<b>0</b>
H. TOTAL DIRECT COSTS (A THROUGH G)							<b>78,431</b>
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
TOTAL INDIRECT COSTS (F&A)							<b>41,569</b>
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)							<b>120,000</b>
K. RESIDUAL FUNDS							<b>0</b>
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)							<b>120,000</b>
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME <b>Bhavani Thuraisingham</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

C \*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET



## BUDGET JUSTIFICATION

**Project P.I.:**  
**Business Address:**

Bhavani Thuraisingham  
Erik Jonsson School of Engineering and Computer Science  
The University of Texas at Dallas, Richardson, TX 75080  
Telephone: (972) 883 - 4738 (voice), 883-2399 (FAX)  
Email: Bhavani.thuraisingham@utdallas.edu  
<http://www.utdallas.edu/~bxt043000/> [www.dr-bhavani.org](http://www.dr-bhavani.org)

### THE UT-DALLAS PROJECT TEAM:

The project team will be led by Dr. Thuraisingham and will include Latifur Khan and Murat Kantarcioglu as well as one part-time graduate assistant. Dr. Thuraisingham will lead in cybersecurity testbed and tool development for the UTD contributions to the infrastructure. She will be assisted by Khan, an active member of the ACM SIGKDD, PAKDD, ICDM and Big Data Analytics communities, and Kantarcioglu whose work focuses on creating technologies that extract useful information from data without sacrificing either security or privacy. They will actively work with the UA team to help integrate their tools into the ISI data infrastructure. The three professors will contribute a total of four and one-half weeks for the entire three year project period. They will be assisted by a part-time graduate student who will provide technology support for system integration (@ 9 months per year for the first two years and 6 months for the final year). Project funding will be directed toward salary and wages, and fringe benefits. Fringe benefit rates are 25% for Thuraisingham, Khan, and Kantarcioglu, and 15% for the graduate student.

### OTHER DIRECT COSTS:

**Travel:** Funding is requested each project year for some travel funding. This will allow one team member each year to actively solicit additional data from the KDD and ICDM communities, assist in seeking community feedback, and promote the infrastructure to other researchers.

### INDIRECT COSTS:

**Indirect cost rate:** 53.0%

The indirect costs have been calculated against the Modified Total Direct Cost base (MTDC). Actual rates in place at the time will be charged to the project funds.

# SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION <b>University of Utah</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Paul Hu</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR				
1. <b>Paul J Hu - PI</b>	0.00	0.00	0.60	<b>2,833</b>			
2.							
3.							
4.							
5.							
6. ( <b>0</b> ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	<b>0</b>			
7. ( <b>1</b> ) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.60	<b>2,833</b>			
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	<b>0</b>			
2. ( <b>0</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00	<b>0</b>			
3. ( <b>1</b> ) GRADUATE STUDENTS				<b>8,250</b>			
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS				<b>0</b>			
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				<b>0</b>			
6. ( <b>0</b> ) OTHER				<b>0</b>			
TOTAL SALARIES AND WAGES (A + B)				<b>11,083</b>			
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				<b>1,708</b>			
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				<b>12,791</b>			
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT				<b>0</b>			
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)				<b>500</b>			
2. FOREIGN				<b>0</b>			
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____				<b>0</b>			
2. TRAVEL _____				<b>0</b>			
3. SUBSISTENCE _____				<b>0</b>			
4. OTHER _____				<b>0</b>			
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> ) TOTAL PARTICIPANT COSTS				<b>0</b>			
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES				<b>0</b>			
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				<b>0</b>			
3. CONSULTANT SERVICES				<b>0</b>			
4. COMPUTER SERVICES				<b>0</b>			
5. SUBAWARDS				<b>0</b>			
6. OTHER				<b>0</b>			
TOTAL OTHER DIRECT COSTS				<b>0</b>			
H. TOTAL DIRECT COSTS (A THROUGH G)				<b>13,291</b>			
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 49.0000, Base: 13292)</b>							
TOTAL INDIRECT COSTS (F&A)				<b>6,513</b>			
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				<b>19,804</b>			
K. RESIDUAL FUNDS				<b>0</b>			
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				<b>19,804</b>			
M. COST SHARING PROPOSED LEVEL \$ <b>0</b> AGREED LEVEL IF DIFFERENT \$							
PI/PD NAME <b>Paul Hu</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

# SUMMARY PROPOSAL BUDGET

YEAR **2**

ORGANIZATION <b>University of Utah</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Paul Hu</b>				AWARD NO.	Proposed	Granted	
				A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)			
				CAL	ACAD	SUMR	
1. <b>Paul J Hu - PI</b>				0.00	0.00	0.60	<b>2,833</b>
2.							
3.							
4.							
5.							
6. ( <b>0</b> ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00	<b>0</b>
7. ( <b>1</b> ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	0.60	<b>2,833</b>
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00	<b>0</b>
2. ( <b>0</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00	<b>0</b>
3. ( <b>1</b> ) GRADUATE STUDENTS							<b>8,250</b>
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS							<b>0</b>
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)							<b>0</b>
6. ( <b>0</b> ) OTHER							<b>0</b>
TOTAL SALARIES AND WAGES (A + B)							<b>11,083</b>
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)							<b>1,708</b>
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)							<b>12,791</b>
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT							<b>0</b>
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							<b>700</b>
2. FOREIGN							<b>0</b>
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____ <b>0</b>							
2. TRAVEL _____ <b>0</b>							
3. SUBSISTENCE _____ <b>0</b>							
4. OTHER _____ <b>0</b>							
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> ) TOTAL PARTICIPANT COSTS							<b>0</b>
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							<b>0</b>
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION							<b>0</b>
3. CONSULTANT SERVICES							<b>0</b>
4. COMPUTER SERVICES							<b>0</b>
5. SUBAWARDS							<b>0</b>
6. OTHER							<b>0</b>
TOTAL OTHER DIRECT COSTS							<b>0</b>
H. TOTAL DIRECT COSTS (A THROUGH G)							<b>13,491</b>
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 49.0000, Base: 13492)</b>							
TOTAL INDIRECT COSTS (F&A)							<b>6,611</b>
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)							<b>20,102</b>
K. RESIDUAL FUNDS							<b>0</b>
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)							<b>20,102</b>
M. COST SHARING PROPOSED LEVEL \$ <b>0</b> AGREED LEVEL IF DIFFERENT \$							
PI/PD NAME <b>Paul Hu</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
				Date Checked	Date Of Rate Sheet	Initials - ORG	

# SUMMARY PROPOSAL BUDGET

YEAR 3

ORGANIZATION <b>University of Utah</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Paul Hu</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
		CAL	ACAD	SUMR			
1.	<b>Paul J Hu - PI</b>	0.00	0.00	0.60	<b>2,833</b>		
2.							
3.							
4.							
5.							
6.	( 0 ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	<b>0</b>		
7.	( 1 ) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.60	<b>2,833</b>		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1.	( 0 ) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	<b>0</b>		
2.	( 0 ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00	<b>0</b>		
3.	( 1 ) GRADUATE STUDENTS				<b>8,250</b>		
4.	( 0 ) UNDERGRADUATE STUDENTS				<b>0</b>		
5.	( 0 ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				<b>0</b>		
6.	( 0 ) OTHER				<b>0</b>		
TOTAL SALARIES AND WAGES (A + B)					<b>11,083</b>		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					<b>1,708</b>		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					<b>12,791</b>		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT					<b>0</b>		
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)					<b>700</b>		
2. FOREIGN					<b>0</b>		
F. PARTICIPANT SUPPORT COSTS							
1.	STIPENDS \$ _____				<b>0</b>		
2.	TRAVEL _____				<b>0</b>		
3.	SUBSISTENCE _____				<b>0</b>		
4.	OTHER _____				<b>0</b>		
TOTAL NUMBER OF PARTICIPANTS ( 0 ) TOTAL PARTICIPANT COSTS					<b>0</b>		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					<b>0</b>		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					<b>0</b>		
3. CONSULTANT SERVICES					<b>0</b>		
4. COMPUTER SERVICES					<b>0</b>		
5. SUBAWARDS					<b>0</b>		
6. OTHER					<b>0</b>		
TOTAL OTHER DIRECT COSTS					<b>0</b>		
H. TOTAL DIRECT COSTS (A THROUGH G)					<b>13,491</b>		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 49.0000, Base: 13492)</b>							
TOTAL INDIRECT COSTS (F&A)					<b>6,611</b>		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					<b>20,102</b>		
K. RESIDUAL FUNDS					<b>0</b>		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					<b>20,102</b>		
M. COST SHARING PROPOSED LEVEL \$ <b>0</b> AGREED LEVEL IF DIFFERENT \$							
PI/PD NAME <b>Paul Hu</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

# SUMMARY PROPOSAL BUDGET Cumulative

ORGANIZATION <b>University of Utah</b>				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Paul Hu</b>				AWARD NO.	Proposed	Granted
					NSF Funded Person-months	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. <b>Paul J Hu - PI</b>				0.00	0.00	1.80
2.						
3.						
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. ( <b>1</b> ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	1.80
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00
2. ( <b>0</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00
3. ( <b>3</b> ) GRADUATE STUDENTS						<b>24,750</b>
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS						<b>0</b>
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						<b>0</b>
6. ( <b>0</b> ) OTHER						<b>0</b>
TOTAL SALARIES AND WAGES (A + B)						<b>33,249</b>
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						<b>5,124</b>
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						<b>38,373</b>
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						<b>0</b>
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						<b>1,900</b>
2. FOREIGN						<b>0</b>
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____ <b>0</b>						
2. TRAVEL _____ <b>0</b>						
3. SUBSISTENCE _____ <b>0</b>						
4. OTHER _____ <b>0</b>						
TOTAL NUMBER OF PARTICIPANTS ( <b>0</b> ) TOTAL PARTICIPANT COSTS						<b>0</b>
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						<b>0</b>
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						<b>0</b>
3. CONSULTANT SERVICES						<b>0</b>
4. COMPUTER SERVICES						<b>0</b>
5. SUBAWARDS						<b>0</b>
6. OTHER						<b>0</b>
TOTAL OTHER DIRECT COSTS						<b>0</b>
H. TOTAL DIRECT COSTS (A THROUGH G)						<b>40,273</b>
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)						
TOTAL INDIRECT COSTS (F&A)						<b>19,735</b>
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						<b>60,008</b>
K. RESIDUAL FUNDS						<b>0</b>
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						<b>60,008</b>
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$		
PI/PD NAME <b>Paul Hu</b>				FOR NSF USE ONLY		
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION		
		Date Checked	Date Of Rate Sheet	Initials - ORG		

C \*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

**University of Utah**

**Budget Justification**

**Title:**        **CI-EN: Computational Research Infrastructure for Intelligence and Security Informatics Research and Community (CRI-ISI)**

**PI:**           **Paul Hu**, Professor, Department of Operations and Information Systems

**I. Personnel**

Paul Hu: Dr. Paul Hu is David Eccles Professor in the Department of Operations and Information Systems, the University of Utah. With research expertise in large-scaled system evaluation, technology acceptance assessment, human-computer interactions, system impact studies, data and text mining, Dr. Hu will lead the system evaluation and assessment efforts; see the preliminary plan in section 5.2 of the Project Description. He will work closely with Ms. Cathy Larson, the University of Arizona project coordinator, to plan and design evaluation studies, and ensure satisfactory implementation of the investigations in a timely fashion. Budgeted time is set at 5% per academic year

Masters Student: Dr. Paul Hu will be assisted by a graduate student, 25% time, who will perform literature search, instrument development and validation, survey designs, experimental task development, pilot study, data collection and coding, and data analyses using statistical and other methods.

**II. Fringe Benefit Rates**

Faculty and Appointed Staff: Hu: 37%  
Graduate Students: 8%

**III. Travel**

A nominal amount of funding is requested to support Dr. Hu's travel for coordination with the project team and direct work with target users.

**IV. Supplies:**

None requested.

**V. Indirect Cost Rate:**

The F&A rate is 49% on all direct costs listed above.

# SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION <b>UNIVERSITY OF VIRGINIA</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Hsinchun Chen</b>				AWARD NO.	Proposed	Granted	
				A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)			
				CAL	ACAD	SUMR	
1. <b>Ahmed Abbasi - PI</b>				0.00	0.00	1.00	<b>20,000</b>
2.							
3.							
4.							
5.							
6. ( 0 ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00	<b>0</b>
7. ( 1 ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	1.00	<b>20,000</b>
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( 0 ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00	<b>0</b>
2. ( 0 ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00	<b>0</b>
3. ( 0 ) GRADUATE STUDENTS							<b>0</b>
4. ( 1 ) UNDERGRADUATE STUDENTS							<b>9,360</b>
5. ( 0 ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)							<b>0</b>
6. ( 0 ) OTHER							<b>0</b>
TOTAL SALARIES AND WAGES (A + B)							<b>29,360</b>
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)							<b>1,200</b>
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)							<b>30,560</b>
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT							<b>0</b>
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							<b>550</b>
2. FOREIGN							<b>550</b>
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____ <b>0</b>							
2. TRAVEL _____ <b>0</b>							
3. SUBSISTENCE _____ <b>0</b>							
4. OTHER _____ <b>0</b>							
TOTAL NUMBER OF PARTICIPANTS ( 0 ) TOTAL PARTICIPANT COSTS							<b>0</b>
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							<b>0</b>
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION							<b>0</b>
3. CONSULTANT SERVICES							<b>0</b>
4. COMPUTER SERVICES							<b>0</b>
5. SUBAWARDS							<b>0</b>
6. OTHER							<b>0</b>
TOTAL OTHER DIRECT COSTS							<b>0</b>
H. TOTAL DIRECT COSTS (A THROUGH G)							<b>31,660</b>
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 58.0000, Base: 31660)</b>							
TOTAL INDIRECT COSTS (F&A)							<b>18,363</b>
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)							<b>50,023</b>
K. RESIDUAL FUNDS							<b>0</b>
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)							<b>50,023</b>
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME <b>Hsinchun Chen</b>				FOR NSF USE ONLY			
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION			
				Date Checked	Date Of Rate Sheet	Initials - ORG	

# SUMMARY PROPOSAL BUDGET

YEAR 2

ORGANIZATION <b>UNIVERSITY OF VIRGINIA</b>				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Hsinchun Chen</b>				AWARD NO.		
				NSF Funded Person-months		Funds Requested By proposer
A. SENIOR PERSONNEL: PI/PI, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. <b>Ahmed Abbasi - PI</b>				0.00	0.00	1.00
2.						
3.						
4.						
5.						
6. ( 0 ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. ( 1 ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	1.00
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. ( 0 ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00
2. ( 0 ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00
3. ( 0 ) GRADUATE STUDENTS						0
4. ( 1 ) UNDERGRADUATE STUDENTS						9,360
5. ( 0 ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						0
6. ( 0 ) OTHER						0
TOTAL SALARIES AND WAGES (A + B)						29,360
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						1,200
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						30,560
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						0
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						550
2. FOREIGN						550
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____				0		
2. TRAVEL _____				0		
3. SUBSISTENCE _____				0		
4. OTHER _____				0		
TOTAL NUMBER OF PARTICIPANTS ( 0 ) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						0
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						0
3. CONSULTANT SERVICES						0
4. COMPUTER SERVICES						0
5. SUBAWARDS						0
6. OTHER						0
TOTAL OTHER DIRECT COSTS						0
H. TOTAL DIRECT COSTS (A THROUGH G)						31,660
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 58.0000, Base: 31660)</b>						
TOTAL INDIRECT COSTS (F&A)						18,363
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						50,023
K. RESIDUAL FUNDS						0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						50,023
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$		
PI/PI NAME <b>Hsinchun Chen</b>  ORG. REP. NAME* <b>Mary Gerrow</b>				FOR NSF USE ONLY		
				INDIRECT COST RATE VERIFICATION		
		Date Checked	Date Of Rate Sheet	Initials - ORG		



# SUMMARY PROPOSAL BUDGET

YEAR 3

ORGANIZATION <b>UNIVERSITY OF VIRGINIA</b>				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Hsinchun Chen</b>				AWARD NO.	Proposed	Granted
				NSF Funded Person-months		
A. SENIOR PERSONNEL: PI/PI, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. <b>Ahmed Abbasi - PI</b>				0.00	0.00	1.00
2.						
3.						
4.						
5.						
6. ( 0 ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. ( 1 ) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	1.00
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. ( 0 ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00
2. ( 0 ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00
3. ( 0 ) GRADUATE STUDENTS						0
4. ( 1 ) UNDERGRADUATE STUDENTS						9,360
5. ( 0 ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						0
6. ( 0 ) OTHER						0
TOTAL SALARIES AND WAGES (A + B)						29,360
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						1,200
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						30,560
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						0
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						550
2. FOREIGN						550
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____				0		
2. TRAVEL _____				0		
3. SUBSISTENCE _____				0		
4. OTHER _____				0		
TOTAL NUMBER OF PARTICIPANTS ( 0 ) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						0
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						0
3. CONSULTANT SERVICES						0
4. COMPUTER SERVICES						0
5. SUBAWARDS						0
6. OTHER						0
TOTAL OTHER DIRECT COSTS						0
H. TOTAL DIRECT COSTS (A THROUGH G)						31,660
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) <b>MTDC (Rate: 58.0000, Base: 31660)</b>						
TOTAL INDIRECT COSTS (F&A)						18,363
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						50,023
K. RESIDUAL FUNDS						0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						50,023
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$		
PI/PI NAME <b>Hsinchun Chen</b>				FOR NSF USE ONLY		
ORG. REP. NAME* <b>Mary Gerrow</b>				INDIRECT COST RATE VERIFICATION		
		Date Checked	Date Of Rate Sheet	Initials - ORG		



## **Budget Justification (University of Virginia)**

### **A. Salaries and Wages – Senior Personnel**

The Principal Investigator, Dr. Ahmed Abbasi, will devote 1 summer month per year throughout the three-year project period. He will work with his fellow collaborators on various facets of the proposed research project. Working closely with the PIs at the University of Arizona and Drexel University, Dr. Abbasi will focus on the phishing website collection. Activities will include managing the research assistants' work regarding test bed construction, refinement of collection and extraction tools, and creation of the phishing website database. Dr. Abbasi will also attend relevant security conferences in order to promote the research test bed and extraction tools made publicly available as a result of the project.

Fringe Benefits are calculated as direct costs in accordance with UVA's indirect cost rate agreement. The university wide rates as of July 1, 2012 are 6.0% for regular 9-month faculty for the three years.

### **B. Salaries and Wages – Other Personnel**

Funding is requested for one research assistants with experience in data and web mining. The research assistants will each work 15 hours/week throughout the year, for the duration of the project. The research assistants will support Dr. Abbasi by performing data collection, processing, phishing database construction, and basic analysis-related activities. They will also support publication and dissemination-related activities. In accordance with UVA policies, the assistants will be paid \$12 per hour.

**C. Permanent Equipment:** Not Applicable.

**D1. Domestic Travel:** For each year, \$1,100 is requested for conference travel. These funds will be used to support Dr. Abbasi and/or students' efforts to present, promote, and disseminate the results of the project, including describing the generated test bed and related extraction tools/resources at security conferences.

**D2. Foreign Travel:** Not Applicable.

**D3. Equipment Purchases:** Not Applicable.

**E. Participant Support Costs:** Not Applicable

**F1. Materials and Supplies:** Not Applicable

**F2. Publication Costs:** Not Applicable

**F3. Consultation Services:** Not Applicable

**F4. Computer (ADP) Services:** Not Applicable

**F5. Subcontracts:** Not Applicable

**F6. Equipment/Facility Rentals:** Not Applicable

**F7. Alterations and Renovations:** Not Applicable

**F8. Other:** Not Applicable.

**H1. Indirect Costs**

Indirect Costs are calculated in accordance with UVA's federally negotiated indirect cost rate agreement, which is currently 58.0% of modified total direct costs (MTDC).

**J. Fee:** Not Applicable

## Current and Pending Support

For: Hsinchun Chen

### Current Support

Taiwan Smart Health Cloud. PI for University of Arizona.

Taiwan National Science Council; \$200,000. 7/1/12-12/31/14. Acad.: .5 mo., Sumr.: .5 mo.

Cybersecurity Scholarship-for-Service at The University of Arizona. PI.

NSF \$4.2M 9/15/2013 – 8/31/2018 Acad. .5 mo

Securing Cyber Space: Understanding the Cyber Attackers and Attacks via Social Media Analytics. PI.

NSF \$1,195,722 9/1/2013 – 8/31/2016 Acad: .36 mo. Sumr: 1.0 mo.

Two Decades of Nanotechnology Development: Global Competitive Landscape and Knowledge Diffusion via ERGM and SIR Analysis. PI.

NSF \$298,358. 10/1/12 – 09/30/14 Sumr: 1.0 mo.

WMD Intent Identification and Interaction Analysis Using the Dark Web

DTRA \$1,047,652 7/14/09 – 07/19/14 Acad: 2.0 mo.

### Pending Support

**[Current Proposal]** CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research and Community (Pilot Demonstration Project). PI.

NSF \$1,499,531 8/1/2014 -7/31/2017 Acad: 0.0 mo.; Sumr: 0.75 mo.

A Systems Approach for Identification and Evaluation of Nanoscience and Nanomanufacturing Opportunities and Risks (EAGER). PI.

NSF \$247,884 6/1/2014 -5/30/2016 Acad: 0.27 mo.; Sumr: 0.45 mo.

Cybersecurity Analytics Education and Training Platform. PI. *Supplement request.*

NSF \$349,848 4/1/14 – 3/30/16 Cal: Yr1: 1.0 mo; Yr2: 0.6 mo.

Securing Cyber Space: Understanding the Cyber Attackers and Attacks via Social Media Analytics. PI. *REU Supplement Request.*

NSF \$56,382 Through 8/31/2016 Sumr. 0.06 mo.

Towards Smart Health and Wellbeing: Information Extraction from Clinical Narratives (III: Small). PI.

NSF \$498,895 7/2014 – 6/2017 Acad: 0.18 mo., Sumr: 0.45 mo.

Toward Smart Chronic Care Using An Intelligent Temporal Analytics Framework (SCH: INT). PI.

NSF \$2,004,471 7/2014 – 6/2018 Acad: 0.10 mo., Sumr: 0.90 mo

## Current and Pending Support

For: Catherine A. Larson

### Current Support

WMD Intent Identification and Interaction Analysis Using the Dark Web. Co-PI (H. Chen, PI).  
DTRA \$1,047,652 7/14/09 – 07/19/14 (NCE) Cal: .0 mo. per project year (ppy)

Cybersecurity Scholarship-for-Service at The University of Arizona. Sr. pers (H. Chen, PI).  
NSF \$2,708,338 8/1/2013 – 7/31/2018 Cal: 1.74 mo ppy

### Pending Support

**[Current Proposal]** CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research and Community (Pilot Demonstration Project). Co-PI. (H. Chen, PI).  
NSF \$1,499,531 8/1/2014 -7/31/2017 Cal: 1.0 mo ppy

A Systems Approach for Identification and Evaluation of Nanoscience and Nanomanufacturing Opportunities and Risks (EAGER). Co-PI. (H. Chen, PI).  
NSF \$247,884 6/1/2014 -5/30/2016 Cal: .60 mo. ppy

Cybersecurity Analytics Education and Training Platform. Senior personnel. Co-PI (H. Chen, PI).  
*Supplement request.*  
NSF \$349,848 4/1/14 – 3/30/16 Cal: Yr1: 1.0 mo; Yr2: 0.6 mo.

## Current and Pending Support

### Mark Patton, Ph.D.

Lecturer, Management Information Systems  
Associate Director, Hoffman E-Commerce Lab  
The University of Arizona

#### Current Support

Cybersecurity Scholarship-for-Service at The University of Arizona.

Role:	Co-PI
Sponsoring Agency:	National Science Foundation
Amount Requested:	\$4,227,303
Start Date:	September 15, 2013
End Date:	August 31, 2018
Time Commitment:	1.7 months per year

#### Pending Support:

CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research and Community (Pilot Demonstration Project). Co-PI.

Agency:	NSF
Amount Being Requested:	\$1,499,531
Proposed Start Date:	August 1, 2014
Proposed End Date:	July 31, 2017
Time Commitment:	1.0 mo per year

## Current and Pending Support

**Ahmed Abbasi, Ph.D.**

Associate Professor of Information Technology  
University of Virginia

### Current Support

EAGER: A User-Centric Approach to the Design of Intelligent Fake Website Detection Systems			
NSF	\$280,172	10/01/2010 – 09/30/2014	Sumr: 0.00
EAGER: Computational Public Drug Surveillance			
NSF	\$155,057	09/01/2012 – 08/31/2014	Sumr: 0.20

### Pending Support

CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research and Community (Pilot Demonstration Project).			
NSF	\$150,068	8/1/2014 -7/31/2017	Sumr: 1.00
SCH: INT: CRUFS – A Unified Framework for Social Media Analysis of Adverse Drug Events			
NSF	\$1,055,072	07/01/2014 – 06/30/2018	Sumr: 0.80
Social Media Analysis of Diversion and Abuse of Buprenorphine and Methadone			
NIH	\$144,126	09/01/2014 – 08/31/2016	Sumr: 0.25



## **Paul Jen-Hwa Hu**

David Eccles Professor of Information Systems

Department of Operations and Information Systems, David Eccles School of Business, University of Utah

Telephone: (801) 587-7785; Fax: (801) 581-7731; Email: paul.hu@business.utah.edu

Personal Web page: [https://faculty.utah.edu/u0239890-Paul\\_Hu/biography/index.html](https://faculty.utah.edu/u0239890-Paul_Hu/biography/index.html)

### **Current and Pending Support**

Current: X. Fang, and P. J. Hu, *Seed Grants*, University of Utah, \$29,000, 2013-2014.

Pending: CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research and Community (Pilot Demonstration Project). Subaward from University of Arizona - PI \$60,008. 2014-2017.





## **Christopher C. Yang, Associate Professor**

**College of Computing and Informatics**

**Drexel University**

**Email: [chris.yang@drexel.edu](mailto:chris.yang@drexel.edu)**

### **Education**

**Ph.D., Electrical and Computer Engineering**, University of Arizona, Tucson, May, 1997.

**M.S., Electrical and Computer Engineering**, University of Arizona, Tucson, May, 1992.

**B.S., Electrical and Computer Engineering**, University of Arizona, Tucson, August, 1990.

### **Work Experience**

**College of Computing and Informatics, Drexel University**

*June 2008 – present: Associate Professor*

**Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong**

*January 2001 – May 2008: Associate Professor*

*January 2000 – December 2000: Assistant Professor*

*January 2000 – May 2008: Director of Digital Library Laboratory*

**Department of Computer Science and Information Systems, University of Hong Kong**

*September 1997 – December 1999: Assistant Professor*

*September 1997 – December 1999: Director of Digital Library Research Laboratory*

*January 1998 – December 1999: Associate Director of Authorized Academic Java<sup>SM</sup> Campus<sup>SM</sup> (AAJC)*

**Selected Publications:** (88 journal papers and 147 conference papers, a complete list can be found at <http://cci.drexel.edu/faculty/cyang/pubs.html>)

(5 related)

1. X. Tang and C. C. Yang, "Detecting Social Media Hidden Communities using Dynamic Stochastic Blockmodel with Temporal Dirichlet Process," *ACM Transactions on Intelligent Systems and Technology*, accepted for publication.
2. Z. Hai, K. Chang, J. Ki, and C. C. Yang, "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance," *IEEE Transactions on Knowledge and Data Engineering*, vol.26, no.3, March 2014, pp.623-634.
3. X. Tang and C. C. Yang, "Ranking User Influence in Healthcare Social Media," *ACM Transactions in Intelligent Systems and Technology*, vol.3, no.4, 2012.
4. C. C. Yang and T. Ng, "Analyzing and Visualizing Web Opinion Development and Social Interactions with Density Based Clustering," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol.41, no.6, 2011, pp.1144-1155.
5. C. C. Yang, X. Shi, and C. Wei, "Discovering Event Evolution Graphs from News Corpora," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol.39, no.4, 2009, pp.850-863.

(5 additional)

6. C. C. Yang and X. Tang, "Estimating User Influence in the MedHelp Social Network," *IEEE Intelligent Systems*, vol.27, no.5, 2012.
7. X. Tang and C. C. Yang, "TUT: A Statistical Model for Detecting Trends, Topics and User Interests in Social Media," *Proceedings of the 21st ACM Conference on Information and Knowledge Management, Maui, October 29 - November 2, 2012.*

8. H. Yang and C. C. Yang, "Harnessing Social Media for Drug-Drug Interactions Detection," *Proceedings of IEEE International Conference on Healthcare Informatics*, Philadelphia, PA, September 8 - 11, 2013.
9. X. Tang, M. Zhang, and C. C. Yang, "Leveraging User Interest to Improve Thread Recommendation in Online Forum," *Proceedings of International Conference on Social Intelligence and Technology*, State College, PA, May 8 – 9, 2013.
10. X. Tang, M. Zhang, and C. C. Yang, "User Interest and Topic Detection for Personalized Recommendation," *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence*, Macau, December 4-7, 2012.

### **Selected Professional Services:**

1. General co-chair, International Conference on Healthcare Informatics 2013, Philadelphia, Pennsylvania (ICHI 2013)
2. General co-chair, International Conference on Social Intelligence and Technology 2013, College Station, Pennsylvania (SOCIETY 2013)
3. Program Committee Chair, ACM SIGHIT International Health Informatics Symposium, Miami, FL, USA, 2012.
4. Program Co-chair, IEEE International Conference on Intelligence and Security Informatics, Vancouver, BC, Canada, 2010.
5. Workshop co-chair, ACM CIKM International Workshop on Smart Health and Wellbeing, Maui, Hawaii (SHB 2012)
6. Workshop Co-chair, ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD), D.C., July, 2010.
7. Track Co-chair, Web 2.0, and Social Media Analytics Track, International Conference on Information Systems (ICIS), Phoenix, AZ, USA, 2009.
8. Program Vice-chair, ACM Eighteenth Conference on Information and Knowledge Management (CIKM), Hong Kong, China, 2009.
9. Guest editor, ACM Transactions on Management Information Systems, Special Issue on Informatics for Smart Health and Wellbeing
10. Guest editor, ACM Transactions on Intelligent Systems and Technology, Special Issues on Intelligent Systems for Health Informatics
11. Guest editor of *IEEE Transactions on Systems, Man, & Cybernetics, Part A*, Special Issue on Social Media Analytics: Understanding the Pulse of the Society.
12. Associate Editor-in-Chief, Security Informatics, Springer
13. Co-Editor, Electronic Commerce Research & Applications, Elsevier

**Collaborators and Co-Editors:** Sophia Ananiadou (University of Manchester, UK), Carlo Combi (Universita' degli Studi di Verona, Italy), Uwe Glaesser (Simon Fraser U., Canada), Wai Lam (Chinese U. of Hong Kong, HK), Ee-Peng Lim (Singapore Management U., Singapore), Jiming Liu (Hong Kong Baptist University), Zhiyong Lu (National Library of Medicine), Yin-Leng Theng (Nanyang Technological University, Singapore), Bhavani Thuraisingham (U. of Texas, Dallas, USA), Jennifer Unger (University of Southern California), Ke Wang (Simon Fraser U., Canada), Chih-Ping Wei (National Tsing Hua U., Taiwan), Flaura Winston (U. of Pennsylvania, USA), John Yen (Pennsylvania State University), Daniel Zeng (U. of Arizona, USA), Qing Zeng (University of Utah)

**Post-Graduate Scholar Sponsor:** Tobun D. Ng, V. Naikan, Shanfeng Zhu.

**Ph.D. and Master Thesis Advisor: Current:** Mi Zhang, Haodong Yang, Ling Jiang, Haemin Kim.  
**Graduated:** Katherine Chuang 12, Xuning Tang 12, Anthony Ng 09, Ying-Chau Wong 08, Jianfeng Lin 07, Nan Liu 06, Xiaodong Shi 06, Tung-Yin Cheng 05, Kar-Wing Li, 04, Patricia Chan 04, Rachel Wong 04, Fu-Lee Wang 03, Milo Yip 02, Mandy Chan 01, Winter Chan 01, Kay Hong 01, Alan Chung 00, Johnny Luk 00, Stanley Yung 00.

## Facilities

### 1. MIS Research Facilities

#### A. Artificial Intelligence Lab at The University of Arizona

The Artificial Intelligence Lab (AI Lab) is an internationally known research group in the areas of digital libraries, intelligent retrieval, collaborative computing, and knowledge management. The group is distinguished for its adaptation and development of scalable and practical artificial intelligence, neural networks, genetic algorithms, statistical analysis, computational linguistics, and visualization techniques. The AI Lab and the Hoffman E-Commerce Lab collaborate on projects and provide access to each other's facilities. Both also have access to facilities provided by the University of Arizona. Dr. Chen is Founding Director of the Hoffman Ecommerce Lab.

- Servers:
  - Dell Poweredge R900, Dell Poweredge 2650, Supermicro SuperBlade, and IntelSC5200 ESX/ESXI Servers as numerous virtual machines for fast data collection
  - Dell Poweredge 1650, Dell Poweredge 2950, Intel SC5600, Lian Li, Supermicro CSE-825TQ, and 3 Supermicro SuperBlade for Project Servers
  - 16-core HPC host with 64 GB Ram; Windows Server 2008
  - Dell 4 cpu Intel Xeon 1.9 GHz processors with 12 GB RAM, Hard Disk storage: 770 GB disk
  - SGI Origin2000 supercomputer, 8 processors and 1 GB RAM, running IRIX (Legacy Platform). 520 GB total storage capacity.
  - 14 other assorted single and dual cpu servers for development, staging, production, and management. 1.8 to 3+ GHz, 1 to 8 GB RAM.
- Server Hosting
  - Dedicated environmentally controlled server room with secured electronic access
- Online Storage
  - 45 TB (30 TB usable) iSCSI storage area network (SAN)
  - 4 TB NAS Box
  - Two 2 TB Direct Attached Storage Units
  - 2 direct attached storage 2 Terabyte storage arrays
  - 1 legacy storage server: Windows 2003 Enterprise Server, 2.6 Terabytes
- Backup and Archival
  - 1 Dell Tape Library
- Workstations
  - Every Lab staff member is provided with one or more workstation(s) fully equipped for their individual needs; specifications vary depending on need. Individuals working on computing-intensive applications are allocated additional workstations and resources depending on need.
  - For special applications, 6 Quad core spidering machines /workstations with quad monitors and 2 TB disk storage are available to Lab members.

#### B. MIS Commons & MicroAge Lab

The MIS Commons & MicroAge Lab is a teaching and research facility originally founded by Dr. Hsinchun Chen and now under the direction of Dr. Mark Patton. This state-of-the-art computer lab is combined with a large multimedia classroom containing an art visual/multimedia presentation system and 26 high-end workstations. Labs and other units part of Eller College have access rights to the resources, which include:

- Servers:
  - 1 Enterprise-class VMware ESX based Xeon virtual server host, 32 GB RAM
  - 2 HP enterprise-class 4 way Windows 2000 Servers, each with 4GB memory

- 1 Enterprise-class IBM/AIX server P660
- 4 AMD 1700+ Secondary servers
- 1 Pentium IV secondary server
- 2 Pentium III secondary servers
- Server Hosting
  - Dedicated environmentally controlled server room with secured electronic access
- Online Storage
  - 1 IBM enterprise-level storage server (2 Terabytes) connected to all Enterprise class servers through a Fiber Channel Storage Area Network (SAN)
- Backup and Archival
  - 1 IBM automated tape storage and retrieval server (126 Terabytes)

## **2. Facilities available through the University of Arizona**

### ▪ **CCIT High Performance Computing Facility**

Personnel also have access to The University of Arizona SGI Altix ICE cluster system which was ranked in June 2008 as the 237th most powerful computer in the world and as the 50th greenest in the world in electricity consumption. The SGI Altix ICE cluster has 1392-core. Each node has an Intel Xeon quad-core processors and a 2GB memory per core. It uses the DDR Infiniband interconnect to form a huge shared memory cluster. This cluster has a computational power of 19.4 TeraFLOPS.

## Data Management Plan

### 1. Background: DIBBs-ISI

This project is intended to create the Data Infrastructure Building Blocks (DIBBs) for the Intelligence and Security (ISI) community of CISE researchers in support of their security-related research. The archive will ultimately be comprised of testbed data containing millions of multilingual social media contents (such as forum messages, tweets, videos, news, and comments) in volatile regions of the world (GeoPolitical Web), cybersecurity related hacker forums, IRC channels, and honeypot information (Hacker Web), and fraudulent phishing web sites (AZPhish Web). The archive will be accessible through an interface that supports searching in meaningful ways, and will provide a variety of tools to support relevant analysis, such as social network analysis, authorship analysis, and data visualization. The DIBBs-ISI cyberinfrastructure will also support collaboration and contributions of new testbeds and tools for use by the entire ISI community.

### 2. Data, Tool, and Testbed Dissemination Activities

Initial project efforts will be aimed at creating a sustainable, web-based DIBBs-ISI research testbed. A database and portal will also be developed and maintained to provide the primary means for researchers accessing the data. Finally, sharing, disseminating, and publicizing the data and tools to appropriate research and other communities will comprise another large part of these efforts.

The Artificial Intelligence (AI) Lab at the University of Arizona already has a strong track record in collecting and sharing data, e.g., the Dark Web collection developed through new deep-web spidering techniques. In data sharing, for example, we are providing graded access to the Dark Web archive to CISE researchers, analysts, investigators and others from different countries and disciplines.

Dissemination and distribution are a major component for the proposed project, for three purposes:

- 1) to ensure that we will gain significant evaluative input from potential users during the data identification phase and during the system development and assessment processes
- 2) to guarantee that such users will have learned about the availability of the DIBBs-ISI testbed and infrastructure for purposes of using it to support their own research and development efforts
- 3) to share our research findings to support the advance of scientific knowledge and discovery in computational sciences.

Efforts are therefore planned for many different kinds of outreach and dissemination, as described here.

*Major ISI and CISE conferences, workshops, and other venues* will be leveraged to promote the DIBBs-ISI archive and infrastructure. For example, the IEEE Intelligence and Security Informatics (ISI) Conference (founded by Dr. Chen; see <http://www.isiconference.org>) provides an excellent outreach opportunity for presenting to the computer and information scientists, security researchers and others who attend. We will also promote our infrastructure in other related ISI meetings in Asia (Pacific-Asia ISI Workshop, PAISI) and Europe (European ISI Conference, EISIC).

The *ACM Knowledge Discovery from Databases (KDD)* community, e.g., through the ISI-KDD workshops and Grand Challenges (organized by Drs. Yang and Chen), will be kept updated about the DIBBs-ISI data and archive availability. Background information about ongoing work with the ACM KDD is in the project description and not repeated here.

*Networking opportunities* will be actively identified. We will also work with ISI and KDD conference organizers to present training sessions on the use of the community resource to interested CISE researchers, and will seek opportunities to participate in other formal and informal “networking” opportunities such as panels, poster sessions, roundtables, etc.

*Organizational publications and listservs* for computational sciences will be utilized to announce availability of the archive, system and tools. Many ACM and IEEE special interest groups (SIGs) also host listservs (for example, for data mining and other specialized topics), to which we have access.

*Formal publication* in computational sciences will be the most important activity for scientific knowledge sharing. The PI has an extremely strong track record in this regard (over 250 peer-reviewed journal articles and 150 refereed conference papers). The Ph.D. students working on the proposed work will also gain extremely valuable experience in writing their results for peer review. Dr. Chen will



promote DIBBs-ISI supported research and other similar projects as the Editor-in-Chief of the ACM *Transactions on Management Information Systems* and the new Springer *Security Informatics* journal.

### **3. Other Considerations**

#### **3.1. Open source data and tools**

The data to be collected (described in proposal text) will generally be open source. Our collection methods will be non-intrusive in that they will not disrupt the normal functioning of the sites being spidered. We will also respect all sources' given use restrictions in order to protect the organization's right to control its own material, and to respect individuals' privacy. All the DIBBs-ISI testbeds are considered open-source contents and will be made freely available to all CISE researchers upon request. Most of our maturing tools will be made available on our portal as open-source components or via collection APIs.

#### **3.2. Human subjects**

Regarding human subjects protection, we have received confirmation from our university's Office of Human Subjects Protection that our proposed use of open-source contents is not considered human subjects research. All system evaluations involving human subjects, however, will have their protocols reviewed and approved through our university's regular HSP processes prior to conducting any research involving human subjects. Based on much previous experience with user evaluations, generally the kinds of studies we describe in the proposal are found to be "exempt," meaning they are exempt from a review by the full Institutional Review Board (IRB), but they must be reviewed and approved at the departmental and university level prior to conducting the research. The process involves providing a complete description of our protocol, including the background, setting, the study population (including vulnerable populations), recruitment and consenting methods, costs or benefits to the subject, procedures used during the study, etc. Additionally, we must also include all materials we will use in communicating with our subjects such as recruitment materials, forms we will use for gaining their informed consent, any written materials such as instructions or questionnaires that subjects will be given, etc.

#### **3.3. Access to DIBBs-ISI**

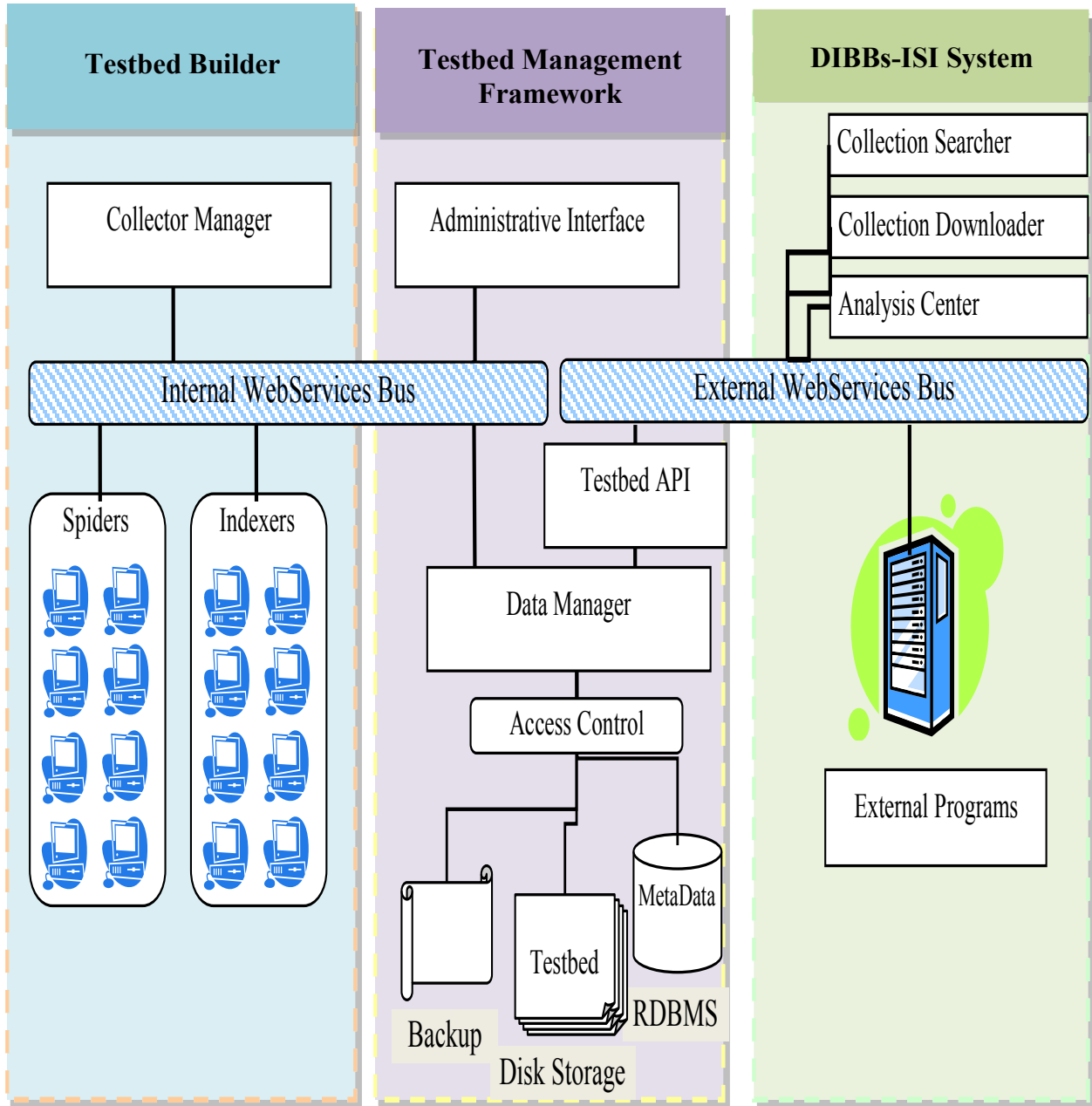
Access to the DIBBs-ISI Portal will be granted to individual users through an online request system. Data about users and their use of the archive and tools may be collected through logs and other means, but individuals' information will not be shared outside of the project team without user assent. Databases, systems, and tools developed will be the intellectual property of the institutions of the contributing researchers. However access and download of DIBBs-ISI data and tools will be free. Post-project maintenance will be provided by the University of Arizona Artificial Intelligence Lab and MicroAge Lab. The AI Lab will provide administrative management and systems support; the MicorAge Lab will provide storage capacity and additional support as needed. The two groups have close working ties and considerable past NSF project experience in the technologies involved to ensure the long-term sustainability and high quality of service after project funding has been expended.

### **4. Monitoring the Outputs and Outcomes of Dissemination Activities**

We will track and report the following: the number of presentations and other sessions we provide, the number and contact information of attendees, and the number who follow up on gaining access; the number and titles of papers published as well as the intended audience and impact factors of the journals; the number and content of listserv messages, ads, and other marketing efforts, as well as the intended audiences. We have experience in doing so in our previously CRI-funded Dark Web project.

This kind of recording and reporting will allow us to remain aware of in which sectors our outreach and publicity efforts are paying off, and which we may need to ramp up. We will also note the sub-disciplines from which people are engaging with the DIBBs-ISI archive. This will allow us to adjust our outreach efforts and target different audiences, conferences, publications, etc. We will also report recommendations and feedback from our planned annual DIBBs-ISI advisory board meeting and comments collected from the various ISI and CISE security-related conferences that we plan to attend.

# DIBBs-ISI System Design Diagram



Dongsong Zhang, Ph.D.  
Associate Professor  
Information Systems Department  
University of Maryland, Baltimore County  
1000 Hilltop Circle  
Baltimore, MD 21250, USA  
Email: zhangd@umbc.edu  
PHONE: 410-455-2851  
FAX: 410-455-1073

March 31, 2014


Prof. Hsinchun Chen  
Artificial Intelligence Lab  
MIS Dept.  
McClelland Hall  
1130 E Helen Street – Rm. 430  
Tucson, AZ 85721 USA

Dear Prof. Chen,

Thank you for the invitation to serve on the Advisory Board for the proposed DIBBs project you are submitting to the National Science Foundation (CIF21 DIBBs: DIBBs for Intelligence and Security Informatics Research and Community [DIBBs-ISI]). I believe the tools and testbed will be of great service to researchers such as myself who are interested in developing and evaluating intelligent information systems by using cutting-edge technologies. I am honored at the invitation to serve on the board and will serve in whatever capacity might be needed. I understand that, as your team develops the infrastructure, you will be seeking feedback and consultation on identifying and selecting appropriate data. We will also be happy to participate in assessing the tools.

I look forward to hearing that the DIBBs-ISI project is funded and hope we may have the opportunity for further collaborations.

Sincerely yours,



Dongsong Zhang, Ph.D.

March 27, 2014

Professor Hsinchun CHEN  
Artificial Intelligence Lab  
MIS Dept., Univ. of Arizona  
1130 E Helen Street  
Tucson, AZ 85721, USA

Dear Hsinchun,

Thank you for inviting me to serve on the Advisory Board for your proposed “Data-Infrastructure Building Blocks” proposal for CISE and ISI researchers (DIBBs-ISI). I gladly accept the invitation and express my interest to advance computational research using this important data and computational testbed.

Social media is no longer new; many of us have already published research results making use of social media data that are in the public domain. Among my own social media research are mining social influence, discovering link structures, modeling social behaviors, mining social media topics and extracting entities and events. Most of these works have been carried out using product review sites, Wikipedia, Question Answering websites, and more recently Twitter.

We have also seen how social media provides unique data to analyse all manner of issues at a scale that could not be performed previously. The big data and deep context in social media also motivates researchers to start collaborating with one another fostering further intelligence and security informatics development and the advanced techniques that support analysis. Unfortunately, even as social media data are available, the huge volume of data and diverse requirements in different research projects often take researchers’ valuable time to write their custom crawlers to gather data before they can start analyzing them.

I therefore fully support the idea of creating a comprehensive testbed to support computational and ISI research. This will definitely expedite the research process and get researchers to focus on data discovery and analytical tasks. I look forward to collaborating with you on this interesting research journey.

Sincerely yours,



Ee-Peng Lim  
*Co-Director, Living Analytics Research Centre*  
*Professor of Information Systems*  
School of Information Systems



**School of Business**  
**The University of Hong Kong**

Pokfulam Road • Hong Kong

Telephone: (852) 2859 1000 • Fax: (852) 2858 5614

Email : enquiry@business.hku.hk

---

香港大學商學院  
香港薄扶林道

March 26, 2014

National Science Foundation  
CISE Directorate  
4201 Wilson Boulevard  
Arlington, VA 22230 U.S.A.

Dear CISE Program Manager:

This letter is to affirm my willingness to serve on the Advisory Board for the Data Infrastructure Building Blocks (“DIBBs”) project proposed by Hsinchun Chen (University of Arizona).

Dr. Chen’s lab excels at data collection, test-bed creation, and tool development. The work proposed would make the current portal even more useful and I wholeheartedly give it my support.

Sincerely,

Michael Chau, Ph.D.  
Associate Professor  
School of Business  
The University of Hong Kong

SCHOOL OF COMPUTING  
QUEEN'S UNIVERSITY

Dr. David B. Skillicorn  
Professor

Goodwin Hall  
Kingston, K7  
Canada  
TEL: +1 613 533 6065  
FAX: +1 613 533 6513  
EMAIL: [skill@cs.queensu.ca](mailto:skill@cs.queensu.ca)  
[www.cs.queensu.ca/home/skill](http://www.cs.queensu.ca/home/skill)

1st April 2014

National Science Foundation  
Directorate - CISE  
4201 Wilson Boulevard  
Arlington, VA 22230 USA

Dear CISE Program Manager:

This letter affirms my willingness to serve on the Advisory Board for Dr. Hsinchun Chen's proposed DIBBs-ISI project, which, if funded, would commence near the end of summer or early Fall 2014.

The team that Dr. Chen has put together for this work is very strong. Each researcher has an excellent record of outreach and publication, and it would be a pleasure to work with them. I hope that NSF will fund the project, as the data and tools proposed will be extremely useful to the CS community.

If I can provide additional information, feel free to contact me.

Yours Sincerely,



Dr. David B. Skillicorn



FACULTY OF APPLIED SCIENCES

March 27, 2014

Prof. Hsinchun Chen  
University of Arizona, MIS Dept.  
McClelland Hall Room 430  
1130 East Helen Street  
Tucson, AZ 85721 USA

Dear Hsinchun;

I am excited to hear that you are proposing to enhance and expand your dark web work. Social media content and discussions continue to be relevant to researchers in the intelligence and security informatics community. Through my work as series editor of Springer Lecture Notes in Social Networks, and as organizer of various conferences related to security and intelligence informatics, such as IEEE ISI 2010 and FOSINT-SI 2014, I clearly see interesting questions and impactful science that can be pursued and answered through your project.

The large, comprehensive testbed to be made available to other researchers would be an invaluable asset. For instance, our experiments in computational criminology, specifically criminal network and organized crime analysis, require massive amounts of data and could greatly benefit from your project. This initiative will help other principal investigators spend more time designing experiments and focusing on analysis, rather than having to spend time on manually collecting data.

Additionally, an infrastructure to help conduct analysis on data would also be of great importance. By having an infrastructure that provides some automated analysis and statistics of data, scientists are able to have some immediate feedback concerning their research.

You have my full support for the proposal you are submitting to the U.S. National Science Foundation, and I am more than happy to be on the advisory board – feel free to contact me with any questions or for input. I look forward to hearing about the success of the project.

Most sincerely,

A handwritten signature in blue ink, appearing to read "U. Glässer".

Dr. Uwe Glässer  
Professor and Associate Dean  
School of Computing Science  
Faculty of Applied Sciences  
Simon Fraser University

OFFICE OF THE DEAN

Tel: 1-778-782-4724  
Fax: 1-778-782-5802  
www.fas.sfu.ca

ADDRESS  
ASB 9861  
Applied Science Building  
8888 University Drive  
Burnaby BC V5A 1S6  
Canada



VirginiaTech

Pamplin College of Business

**Business Information Technology**

1007 Pamplin Hall (0235), Blacksburg, Virginia 24061

540/231-5074 Fax: 540/231-3752

E-mail: alanwang@vt.edu

[www.wang.bit.vt.edu](http://www.wang.bit.vt.edu)

March 28, 2014

National Science Foundation

Directorate for Computer & Information Science and Engineering

4201 Wilson Blvd.

Arlington, VA 22230

Dear Program Manager:

This letter of support is for Dr. Hsinchun Chen's proposed project, "DIBBs for Intelligence and Security Informatics Research and Community," and affirms my great willingness to serve on the advisory board he has put together to provide an outside perspective.

I would personally be extremely gratified to see the Dark Web Forum Portal expanded and enhanced to include a broader testbed of data. I believe the data and tools will both be supremely useful to supporting research in information and computer sciences and hope to see this project funded.

Sincerely,

G. Alan Wang, Ph.D.

Associate Professor in Business Information Technology

Pamplin College of Business

Virginia Tech

*Invent the Future*

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

*An equal opportunity, affirmative action institution*