



AZSecure-data.org

Spring Data Release

Data Infrastructure Buildings Blocks for Intelligence and Security Informatics (DIBBs for ISI)

Thank you to those who attended the Cybersecurity Big Data and Analytics Sharing Workshop at the WiCyS conference in Tucson this March, it was a pleasure meeting you. The [workshop report](#) and [presenters' slides](#) are now available.

The IEEE ICDM Workshop on Intelligence and Security Informatics 2017 (ISI-ICDM 2017) will be held on Nov 18. Details can be found at <http://cci.drexel.edu/isi/isi-icdm2017/>.

We hope to see you at [IEEE-ISI 2017](#) in Beijing, China July 22-24!

New Data

The AZSecure-data.org data portal has released new intelligence and security informatics-related data sets. Security researchers are encouraged to explore the collections and announce their availability to their students and other researchers.

The newly added data sets include the following:

- [ADFA IDS 2017](#) - this is an update to the original ADFA IDS dataset.
- [CSDMC 2010](#) is a historical reference collection of malware traces collected during a data mining competition in 2010. This dataset contains both classified and unclassified logs for training purposes.
- [Twitter Identity](#) is a collection of 976 randomly selected Twitter histories collected between October 2014 and February 2015. Twitter data is available upon request.
- [Wi-Fi Header Database](#) contains 9,817,671 wi-fi traffic headers collected between 6/3/2016 – 6/11/2016.
- An in-progress collection of phishing websites joined our previous phishing collections in September. [PhishMonger](#) collects phishing websites as they are reported and before these ephemeral websites can disappear. The University of Virginia researchers David Dobolyi and Ahmed Abbasi are making this collection available as they compile it. A third batch of data was added to AZSecure-data.org in April. As of output #8409, the collection contains ~252,000 websites collected between November 2015 and April 2017. The PhishMonger collection can be used for the development or improvement of anti-phishing tools, for usability studies and susceptibility experiments, as well as time-series and longitudinal analysis of phishing attacks and malware. New data is now added monthly.



Hsinchun Chen, Ph.D.

Regents' Professor,
Thomas R. Brown Chair of
Management and
Technology, and Director
of the Artificial
Intelligence Lab,
University of Arizona.

PI and Director for Data
Infrastructure Building
Blocks for Intelligence and
Security Informatics
(DIBBs for ISI).

hchen@eller.arizona.edu

University of Arizona
Tucson, AZ 85721 USA
520.621.2748 tel
520.521.2433 fax

ai.arizona.edu

New Tools Section

AZSecure-data also hosts [tools and tutorials](#). Check out the new Inventory of ISI Research and Analysis Tools. The inventory provides a summary of tools that can aid researchers in performing data driven security analytics based on the tools used in the University of Arizona's Artificial Intelligence Lab. The inventory is divided into three major sections based on a traditional data analytics pipeline: (1) collection and storage tools; (2) pre-processing and analytics tools; and (3) visualization tools. For each category, a short summary of the typical types of tasks that are completed in that phase of the data analytics procedure is given, followed by an inventory of tools that fall into that category. The name of each tool is accompanied by a link of where to download and get documentation.

Resources such as this are of great utility to students and others learning how to perform research functions. AZSecure-data is an excellent place to share tools and tutorials with the research community. If you have any tools or tutorials to add, please contact us at ailab@eller.email.com.

Preparing to Migrate to DSpace

DSpace is a type of content management system specifically designed for archival use. It allows for many functions that the current test portal does not. It gives granular control of metadata, so users can search for authors and keywords or browse collections that are listed in multiple places based on their topics or contents. DSpace automatically generates XML metadata in compliance with OAI-PMH. This allows the metadata in our system to be collected by other research indexes and breaks the barriers of siloed data. Persistent identifiers will allow easier citation and access to datasets the same way DOIs work for journal articles. DSpace also allows users to subscribe to specific topics of interest and receive notification when new materials are added to those topics.

Migration to DSpace should be completed in time for the Summer Newsletter and the beginning of the Fall 2017 semester.

Share Your Data

AZSecure-data continues to accept legacy and current datasets from the intelligence and security informatics domain. You can recommend or submit a dataset at any time by contacting ailab@eller.arizona.edu. **AZSecure-data.org is an excellent solution for funder-mandated data sharing.**

About the Project

With funding from the National Science Foundation, this Data Infrastructure Building Blocks for Intelligence and Security Informatics ("DIBBs for ISI") project is developing a research infrastructure for use by ISI scientists, students, and others studying a wide range of cyber- and security-related issues, whether in the computer, information, or social sciences. The project portal, AZSecure-data, provides researchers and students with free access to a wide variety of relevant data sets. This material is based upon work supported by the National Science Foundation Grant No. ACI-1443019.

In January Dr. Chen and Cathy Larson presented on the DIBBs-ISI project to NSF and other DIBBs PIs at the PI meeting. The associated [slides](#), [poster](#), and [white paper](#) are now available on the DIBBs Portal.

Interested in learning more?

Join our Intelligence and Security Informatics Group on [LinkedIn](#), [Facebook](#), and [Twitter](#).

Do you have data to share?

We can help! Send an inquiry to ailab@email.arizona.edu. If you know someone privately hosting or looking to host research data, send them our way!

Tell Us Your Data Story!

Data stories help convey the importance of making sharable data easy to locate and reuse. Tell us your story about working with the datasets we host: which dataset(s) you used and how you used them, what you were investigating, and what you found out.

Staff Changes at AZSecure-data

Cathy Larson retired from the AI Lab after working with Dr. Chen for 14 years. Please join us in wishing her a restful retirement.

This material is based upon work supported by the National Science Foundation under Grant No. ACI-1443019.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the

Follow Us!

This newsletter is quarterly, but if you desire more frequent updates on ISI related topics, take a look at Intelligence and Security Informatics on [LinkedIn](#), [Facebook](#), and [Twitter](#). If you missed a newsletter, you can always [read them on the portal](#).



To be removed from this distribution list, click [here](#) or reply with REMOVE in the subject line:

##OptOutURLSingleList##

Artificial Intelligence Laboratory
The University of Arizona
1130 E. Helen Street
McClelland Hall 116
Tucson, AZ 85721

##TrackImpression##