



<https://ai.arizona.edu/research/dibbs>

AZSecure-data.org

PI: Dr. Hsinchun Chen, University of Arizona. Co-PIs: Dr. Mark Patton and Cathy Larson, University of Arizona. Project Partners: Dr. Ahmed Abbasi, University of Virginia; Dr. Paul Hu, University of Utah; Dr. Bhavani Thurasingham, University of Texas at Dallas; Dr. Chris Yang, Drexel University.

What is ISI? ISI the field for researching and developing advanced information technologies and systems for national and international security-related applications, through an integrated technological, organizational, and policy-based approach.

Who is ISI? The community of ISI researchers has emerged since the inaugural ISI Workshop sponsored by NSF, DHS, and DOD in 2003, and is now made up of an estimated 4,000+ international scholars, scientists, students, and others engaged in the field.

What is DIBBs for ISI? DIBBs for ISI Research and Community is our project to create community infrastructure through a portal of ISI data and tools to help build and serve the national and international intelligence and security informatics research communities.

Project Successes

Data Science Testbed for Security Researchers

From the repository, researchers can access and download zipped data files and their metadata. See details at right for information on available data. Access the Dark Web and Geo Web forums, Phishing websites, and more.

AZSecure-data.org
Intelligence and Security Informatics Data Sets

Data Infrastructure Building Blocks for ISI. A Project of the University of Arizona (NSF #ACI-1443019), Drexel University, University of Virginia, University of Texas at Dallas, and University of Utah

Home | About | Get Data | ISI Events | Citing Data | Papers | Policies

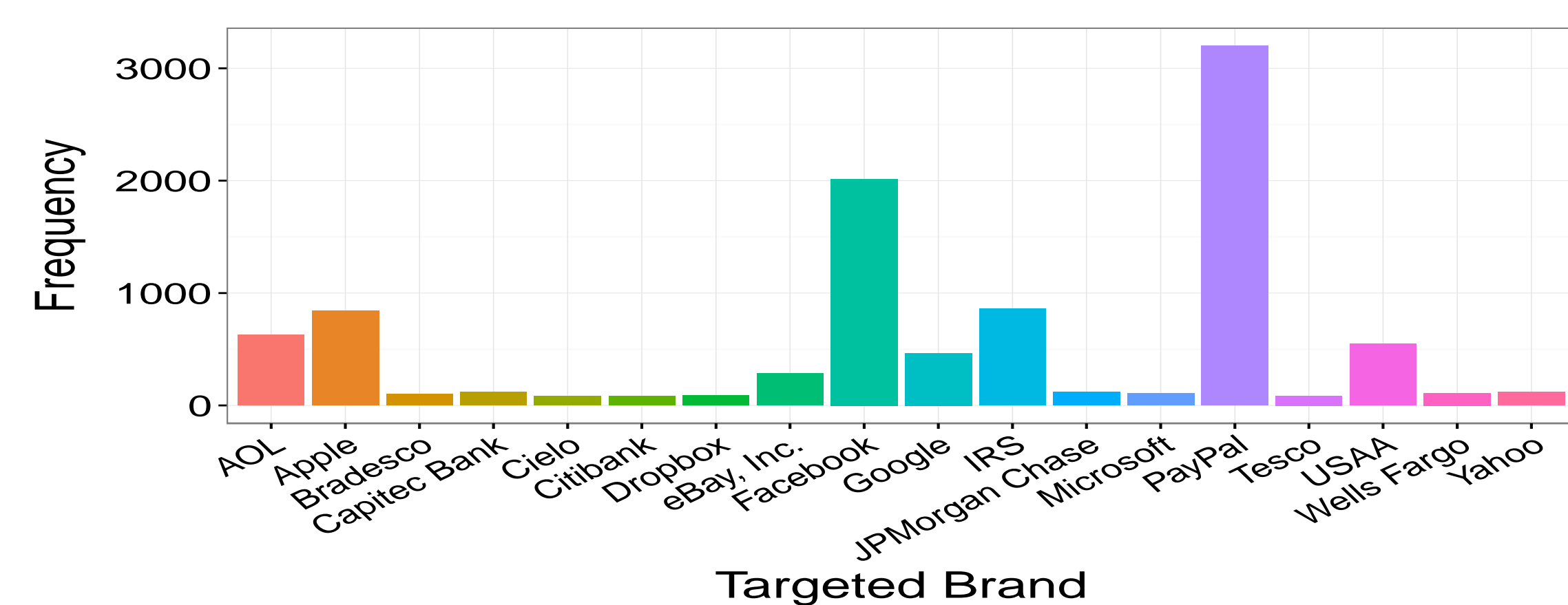
Get Data

- DARK WEB FORUMS
- GEOWEB FORUMS
- PHISHING SITES
- TWITTER DATA
- OTHER FORUMS
- OTHER DATA

The AZSecure-data repository currently provides access to Web forums, Internet phishing websites, Twitter data, and other data. Most files are available to download from the "Get Data" buttons above; other files can be requested through the project manager. To request access to restricted data, send email to aialab@eller.arizona.edu and state which data set you would like to use and the purpose for which it will be used; also provide complete contact information including name, affiliation and mailing address, email, and telephone number.

Tools: AZPhishmonger

AZPhish Web's PhishMonger Module, developed for the DIBBs project, collects large-scale phishing websites while they are live on the web: over 170,000 collected to date!



Frequency of phishing websites for targeted brands that contain a minimum of 80 instances while excluding the category "Other." Financial services are a common phishing target.

The ISI Data Repository: AZSecure-data.org

Shared datasets allow researchers to test algorithms and models against the results and methods of others, and provide a wealth of security informatics data in a timely manner, allowing them to focus on new research. Important collections include the following:

- THE WORLD'S LARGEST COLLECTION OF PHISHING WEBSITES:** Use this collection - the most popular in the portal - of over 170,000 phishing websites to analyze phishing attack trends by industry and brand, over time, as well as attack volume forecasting and time-series modeling.

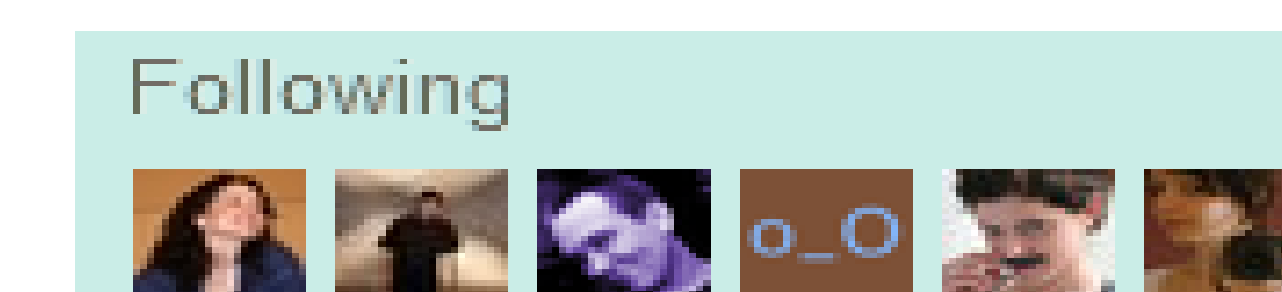


- OVER 39,000,000 MILLION EXTREMIST AND SOCIAL MOVEMENT POSTINGS FROM 93 UNIQUE FORUMS,** from the AI Lab's Dark Web & Geo Web projects, in English, Arabic, French, German, Pashto, Russian, and Urdu: Study extremism and other state and non-state social movements and communications in at-risk regions.

- HACKER COMMUNITY DATA - DARKNET MARKETPLACES, HACKER FORUMS, IRC CHANNELS, BITCOIN TRANSACTIONS, AND CARDING SHOPS [forthcoming]:** A selection from more than 185,000,000 records from 79 platforms, collected as part of our NSF-funded Hacker Web project, will be added to the AZSecure-data repository for sharing with the ISI community to study the hacker community and behavior.

- PATRIOT, MILITIA, HATE, AND LINKED WEBSITES:** Hundreds of historical websites the Southern Poverty Law Center identified as promoting extreme social perspectives: research white supremacy/extremism, and linkages between virtual and real organizations.

- EVEN MORE!** Twitter data, malware and analysis data, modern network traffic data: Study malware, black markets, hackers, dark social networks, and more.



- Social media is often ephemeral. Our collections contain materials that are no longer available in their original contexts. These become valuable historical records for digital humanities research, supporting cross-disciplinary reuse. In the most recent quarter, portal users have downloaded 269GB of data, with 3,020 requests initiated from 22 different countries and regions.

Community Building

In addition to providing the repository of ISI data and tools, the DIBBs for ISI project promotes collaboration and community by:

- Reaching out frequently to over 2,300 potential ISI users and contributors
- Evaluating the repository through regular feedback from users
- Organizing data challenges in the ASONAM, ICDM, ISI, and KDD conferences. These conferences are attended annually by over 3,000 researchers and students!
- Organizing and sponsoring workshops in the upcoming 2017 Women in Cybersecurity conference (800+ attendees) and IEEE ISI Conference (200+ attendees)

Join Us

