THE UNIVERSITY OF ARIZONA

Data Infrastructure Building Blocks (DIBBs) for Intelligence and Security Informatics Research and Community

- Identifying and Building the ISI Community
- They Want Our Data! Developing the Data Science Testbed
- Promoting Collaboration, Education, and Community
- Demonstration, Plans, and Future Growth
- Join Us!

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.

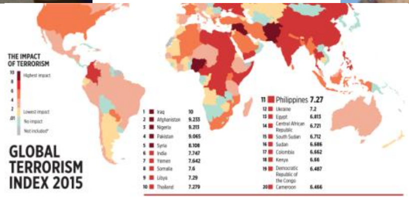
This material is based in part upon work supported by the National Science Foundation under Grant Number ACI-1443019.

Building the ISI Community

- ISI = "**Intelligence and Security Informatics**"
- ISI = R&D for **advanced information technologies** and systems for **national and international security-related applications**
  - Through an integrated technological, organizational, and policy-based approach
- ISI community first brought together at the inaugural ***Intelligence and Security Informatics International Conference***
  - Initially founded by Dr. Hsinchun Chen (UA), sponsored by the **National Science Foundation** and held in Tucson in 2003
  - Now under the auspices of IEEE with multiple international events and a community of 4,000+ scholars in US, Europe, and Asia

2008 IEEE International Conference on
Intelligence and Security Informatics
June 17-20 2008, Taipei, Taiwan
IEEE

Why ISI?



Since the year 2000, there have been over 73,000 terrorist attacks, killing more than 170,000 people.

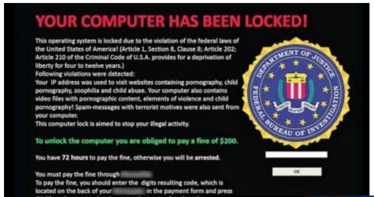
The global economic impact of terrorism reached US$89.6 billion in 2015.

-- Global Terrorism Index, 2015

Cyber crime costs are estimated to be $400 billion a year, with estimates going as high as $2 trillion by 2019.

Large banks, retailers, and federal agencies make the headllines, but all businesses are at risk.

-- Forbes, US edition, 1/17/2016



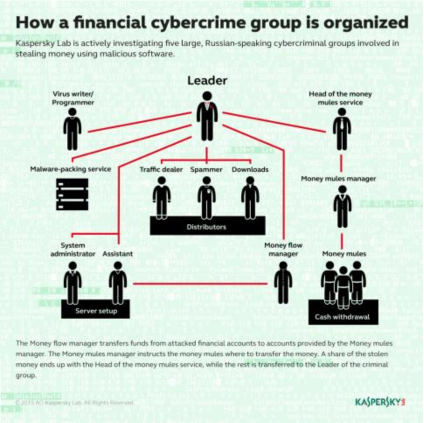
YOUR COMPUTER HAS BEEN LOCKED!

This operating system is locked due to the violation of the federal laws of the United States of America! (Article 1, Section 8, Clause 8; Article 202; Article 210 of the Criminal Code of U.S.A. provides for a deprivation of liberty for four to twelve years.)
Following violations were detected:
Your IP address was used to visit websites containing pornography, child pornography, zoophilia and child abuse. Your computer also contains video files with pornographic content, elements of violence and child pornography! Spam-messages with terrorist motives were also sent from your computer.
This computer lock is aimed to stop your illegal activity.

To unlock the computer you are obliged to pay a fine of $200.

You have **72 hours** to pay the fine, otherwise you will be arrested.

You must pay the fine through [redacted]
To pay the fine, you should enter the digits resulting code, which is located on the back of your [redacted] in the payment form and press OK (if you have several codes, enter them one after the other a
OK).



How a financial cybercrime group is organized

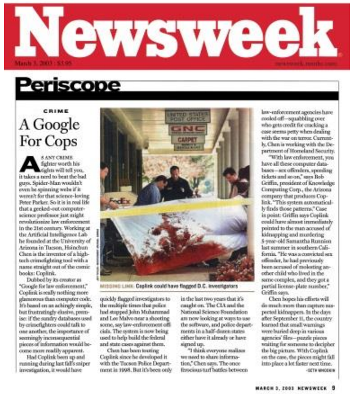
Kaspersky Lab is actively investigating five large, Russian-speaking cybercriminal groups involved in stealing money using malicious software.

The Money flow manager transfers funds from attacked financial accounts to accounts provided by the Money mules manager. The Money mules manager instructs the money mules where to transfer the money. A share of the stolen money ends up with the Head of the money mules service, while the rest is transferred to the Leader of the criminal group.

They Want Our Data!

A Short Course in AI Lab **ISI Data History**:

– **"A Google for Cops"**: With funding from NIH, NSF, etc., the Lab develops **COPLINK** to support **information sharing, analysis, and visualization of law enforcement data**; COPLINK was commercialized and purchased by IBM and is now used by 3,500+ agencies. New approaches in data/text/web mining in COPLINK lead to Dark Web:



Newsweek

March 3, 2003 : $3.95

newsweek.msnbc.com

Periscope

CRIME

A Google For Cops

As any crime fighter worth his sights will tell you, it takes a nerd to beat the bad guys. Spider-Man wouldn't even be spinning webs if it weren't for that science-loving Peter Parker. So it is in real life that a geeked-out computer-science professor just might revolutionize law enforcement in the 21st century. Working at the Artificial Intelligence Lab he founded at the University of Arizona in Tucson, Hsinchun Chen is the inventor of a high-tech crimefighting tool with a name straight out of the comic books: Coplink.

Dubbed by its creator as "Google for law enforcement," Coplink is really nothing more glamorous than computer code. It's based on an achingly simple, but frustratingly elusive, premise: if the sundry databases used by crimefighters could talk to one another, the importance of seemingly inconsequential pieces of information would become more readily apparent.

Had Coplink been up and running during last fall's sniper investigation, it would have quickly flagged investigators to the multiple times that police had stopped John Muhammad and Lee Malvo near a shooting scene, say law-enforcement officials. The system is now being used to help build the federal and state cases against them.

Chen has been touting Coplink since he developed it with the Tucson Police Department in 1998. But it's been only in the last two years that it's caught on. The CIA and the National Science Foundation are now looking at ways to use the software, and police departments in a half-dozen states either have it already or have signed up.

"I think everyone realizes we need to share information," Chen says. The once ferocious turf battles between law-enforcement agencies have cooled off—squabbling over who gets credit for cracking a case seems petty when dealing with the war on terror. Currently, Chen is working with the Department of Homeland Security.

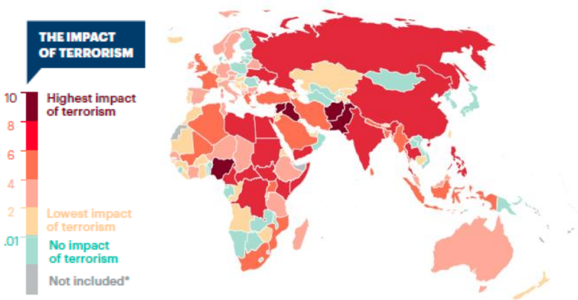
"With law enforcement, you have all these computer databases—sex offenders, speeding tickets and so on," says Bob Griffin, president of Knowledge Computing Corp., the Arizona company that produces Coplink. "This system automatically finds those patterns." Case in point: Griffin says Coplink could have almost immediately pointed to the man accused of kidnapping and murdering 5-year-old Samantha Runnion last summer in southern California. "He was a convicted sex offender, he had previously been accused of molesting another child who lived in the same complex, and they got a partial license-plate number," Griffin says.

Chen hopes his efforts will do much more than capture suspected kidnappers. In the days after September 11, the country learned that small warnings were buried deep in various agencies' files—puzzle pieces waiting for someone to decipher the big picture. With Coplink on the case, the pieces might fall into place a lot faster next time.

—SETH MNOOKIN

MISSING LINK: Coplink could have flagged D.C. investigators

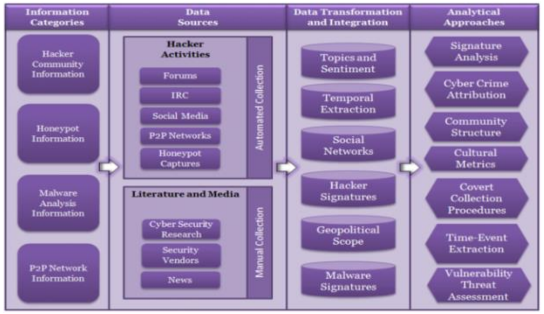
MARCH 3, 2003 NEWSWEEK 9



From: Global Terrorism Index, 2015

- The **Dark Web:** terrorists' and cyber criminals' use of the Internet. Generates **massive amounts of in-demand data: forums, websites, videos, blogs, etc.** From Dark Web to Hacker Web (cybersecurity) via **new approaches to spidering, social network analysis, authorship analysis, video analysis, sentiment analysis**, and more.

– The **HackerWeb** project uses **security big data analytics** to answer important questions about **hacker behaviors, markets, community structure, communication, and cultural differences**. (Research framework at right.)



Information Categories: Hacker Community Information; Honeypot Information; Malware Analysis Information; P2P Network Information

Data Sources: Hacker Activities (Automated Collection): Forums; IRC; Social Media; P2P Networks; Honeypot Captures. Literature and Media (Manual Collection): Cyber Security Research; Security Vendors; News

Data Transformation and Integration: Topics and Sentiment; Temporal Extraction; Social Networks; Hacker Signatures; Geopolitical Scope; Malware Signatures

Analytical Approaches: Signature Analysis; Cyber Crime Attribution; Community Structure; Cultural Metrics; Covert Collection Procedures; Time-Event Extraction; Vulnerability Threat Assessment

BIG DATA + SECURITY NEEDS + RESEARCHERS + NSF = DATA INFRASTRUCTURE BUILDING BLOCKS FOR INTELLIGENCE AND SECURITY INFORMATICS

AZSecure-data.org: Developing the Data Science Testbed

Emphasize difficult to collect, novel data sources:

- **Dark Web Data** and **Geo Web Data**: the most **in-demand** and **easiest/fastest to make available**
  - Originally requested by over 800 researchers, students, analysts, others, in computer science, international relations, intelligence
  - **39M+ postings** from general, extremist, and terrorist forums in their original languages
- **PhishMonger**: A **new tool** developed as part of DIBBs to collect **live phishing websites** to support the study and analysis of these ephemeral exploits[1]
  - Currently the most requested data: **227 GB** downloaded in the recent period (September-November)

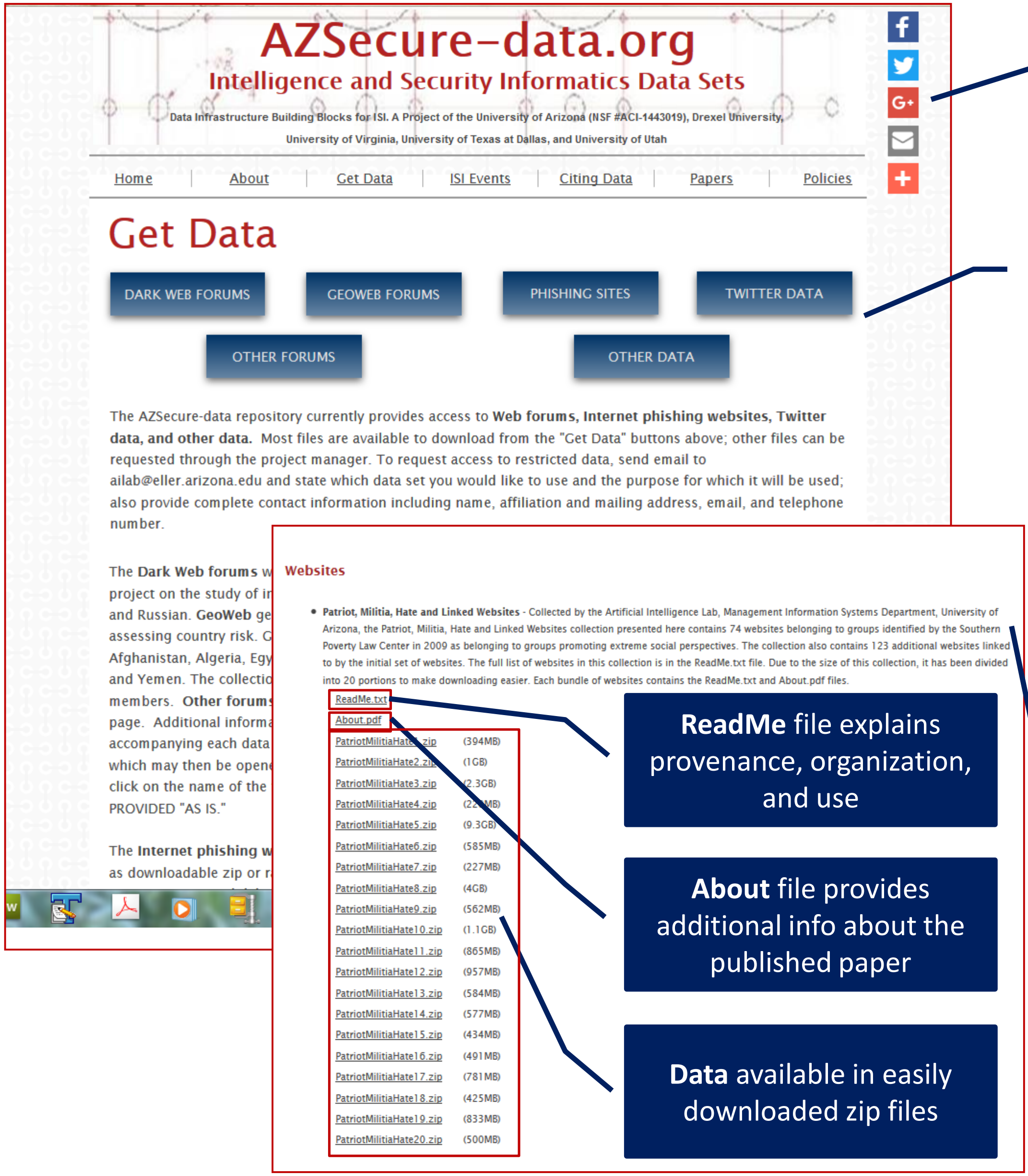
[1] For system details, see Dobolyi & Abbasi, 2016. “PhishMonger: A Free and Open Source Public Archive of Real-World Phishing Websites,” *in* Proceedings of the IEEE ISI 2016 Conference.

Promoting Collaboration, Education, and Community

Build, promote, and support the ISI community

- **Data Challenges:** Organized around a question or series of questions which can only be answered by the use of the AZSecure-data set provided.
  - Encourages **new analytical techniques** and **promotes the use of the portal**
  - Successful data challenges @ **IEEE's ICDM** (International Conference on Data Mining), **ASONAM** (Advances in Social Network Analysis and Mining), **KDD** (ACM's Knowledge Discovery and Data Mining conference), reaching a potential audience of more than 3,000 attendees
- **New Workshops:** Two **Security Big Data and Analytics Sharing Workshops** will be held in 2017 and 2018 to get user input to build on the DIBBs-supported platform
- **Quarterly Updates:** Sent to list of **2,300+ potential ISI users**, compiled from ISI proceedings and directories of institutes, centers, labs, and other security related organizations

Demonstration, Plans, and Future Growth



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

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 ailab@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.

The **Dark Web forums** w
project on the study of ir
and Russian. **GeoWeb** ge
assessing country risk. G
Afghanistan, Algeria, Egy
and Yemen. The collectio
members. **Other forums**
page. Additional informa
accompanying each data
which may then be open
click on the name of the
PROVIDED "AS IS."

The **Internet phishing w**
as downloadable zip or r

Websites

- **Patriot, Militia, Hate and Linked Websites** - Collected by the Artificial Intelligence Lab, Management Information Systems Department, University of Arizona, the Patriot, Militia, Hate and Linked Websites collection presented here contains 74 websites belonging to groups identified by the Southern Poverty Law Center in 2009 as belonging to groups promoting extreme social perspectives. The collection also contains 123 additional websites linked to by the initial set of websites. The full list of websites in this collection is in the ReadMe.txt file. Due to the size of this collection, it has been divided into 20 portions to make downloading easier. Each bundle of websites contains the ReadMe.txt and About.pdf files.

ReadMe.txt
About.pdf

PatriotMilitiaHate1.zip	(394MB)
PatriotMilitiaHate2.zip	(1GB)
PatriotMilitiaHate3.zip	(2.3GB)
PatriotMilitiaHate4.zip	(22[illegible]MB)
PatriotMilitiaHate5.zip	(9.3GB)
PatriotMilitiaHate6.zip	(585MB)
PatriotMilitiaHate7.zip	(227MB)
PatriotMilitiaHate8.zip	(4GB)
PatriotMilitiaHate9.zip	(562MB)
PatriotMilitiaHate10.zip	(1.1GB)
PatriotMilitiaHate11.zip	(865MB)
PatriotMilitiaHate12.zip	(957MB)
PatriotMilitiaHate13.zip	(584MB)
PatriotMilitiaHate14.zip	(577MB)
PatriotMilitiaHate15.zip	(434MB)
PatriotMilitiaHate16.zip	(491MB)
PatriotMilitiaHate17.zip	(781MB)
PatriotMilitiaHate18.zip	(425MB)
PatriotMilitiaHate19.zip	(833MB)
PatriotMilitiaHate20.zip	(500MB)

ReadMe file explains provenance, organization, and use

About file provides additional info about the published paper

Data available in easily downloaded zip files

Social sharing links

Data types currently available

- Dark Web and Geo Web forums
- Phishing sites
- Twitter data
- Other forums
- Other data:
  - Malware
  - Network traffic
  - News
  - Web chat
  - Websites

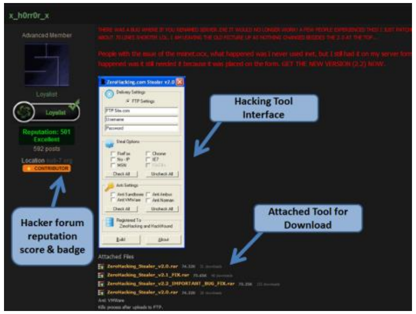
EXAMPLE: Patriot, Militia, Hate, & Linked Websites

Demonstration, Plans, and Future Growth

Coming Soon!

1) **Hacker Community Data** - a selection of more than **185,000,000 records** from **79 platforms** to **study the international hacker community and behavior**, including:

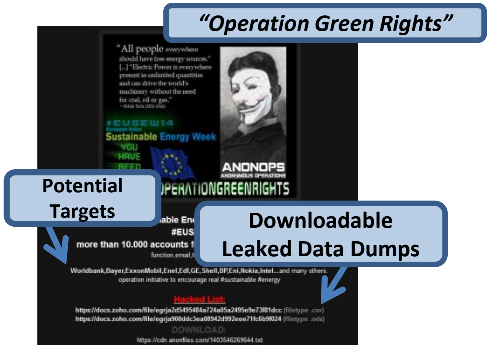
   - Darknet Marketplaces
   - Hacker Forums
   - IRC Channels
   - Bitcoin transactions
   - Carding shops



People with the issue of the mainet.ocx, what happened was I never used mainet, but I still had it on my server form [illegible] happened was it still needed it because it was placed on the form. GET THE NEW VERSION (2.2) NOW.

ZeroHacking_Stealer_v2.0.rar
ZeroHacking_Stealer_v2.1_FIX.rar
ZeroHacking_Stealer_v2.2_IMPORTANT_BUG_FIX.rar
ZeroHacking_Stealer_v2.0.rar
Anti VMWare
Kills process after uploads to FTP.

Figure 1. Hacker forum posting where a member shares malicious credential stealing tool.



"All people everywhere should have free energy sources." [...] "Electric Power is everywhere present in unlimited quantities and can drive the world's machinery without the need for coal, oil or gas."
– Nikola Tesla

more than 10.000 accounts f
Worldbank,Bayer,ExxonMobil,Enel,Edf,GE,Shell,BP,Eni,Nokia,Intel....and many others.
operation initiative to encourage real #sustainable #energy

Hacked List:
https://docs.zoho.com/file/egrja2d5495484a724a05a2495e9e73f81dcc (filetype .csv)
https://docs.zoho.com/file/egrja900ddc3ea08942d992eee71fc6b9f024 (filetype .ods)
DOWNLOAD:
https://cdn.anonfiles.com/1403546269644.txt

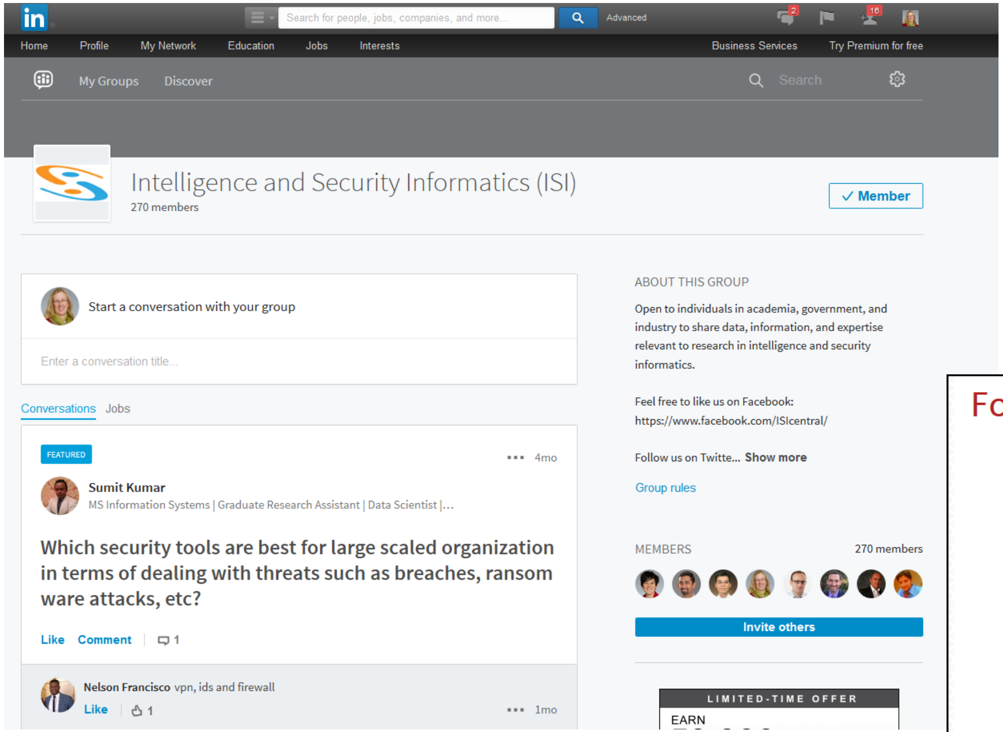
Figure 2. Dumps of compromised credentials in a hacker forum posting.

2) Dunning **ISI researchers** to **share their data**!

   - Cold-calling and otherwise contacting our community of 4,000+ researchers to prepare and share their data via the AZSecure-data portal.

3) New **repository platform** to provide a stable environment with increased searching, browsing, and analytical functionality

Join Us!



Intelligence and Security Informatics (ISI)

270 members

Start a conversation with your group

Enter a conversation title...

Sumit Kumar 4mo

MS Information Systems | Graduate Research Assistant | Data Scientist |...

Which security tools are best for large scaled organization in terms of dealing with threats such as breaches, ransom ware attacks, etc?

Like Comment 1

Nelson Francisco vpn, ids and firewall

Like 1 1mo

ABOUT THIS GROUP

Open to individuals in academia, government, and industry to share data, information, and expertise relevant to research in intelligence and security informatics.

Feel free to like us on Facebook: https://www.facebook.com/ISIcentral/

Follow us on Twitte... Show more

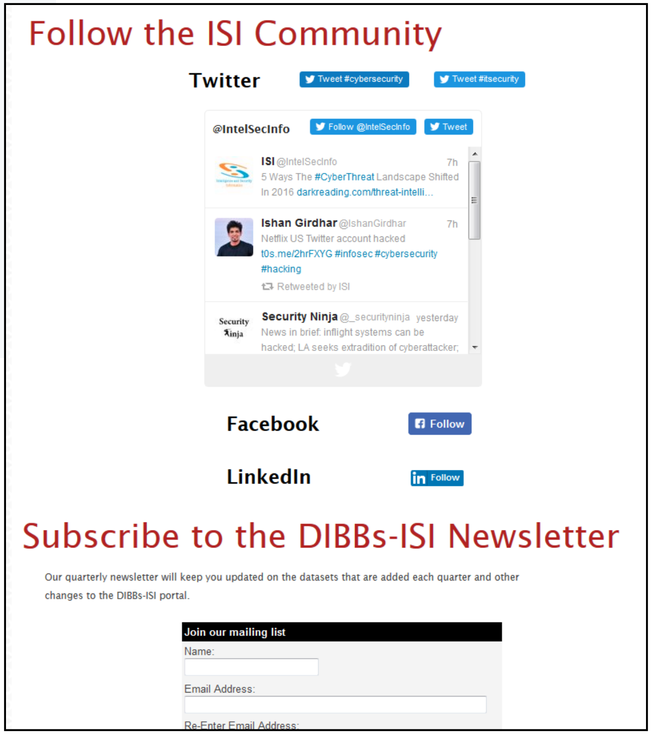
Group rules

MEMBERS 270 members

LIMITED-TIME OFFER

EARN

- 270 members in LinkedIn ISI group
- Building up Facebook and Twitter activity
- Sending newsletter quarterly



Follow the ISI Community

Twitter

@IntelSecInfo

ISI @IntelSecInfo 7h
5 Ways The #CyberThreat Landscape Shifted In 2016 darkreading.com/threat-intelli...

Ishan Girdhar @IshanGirdhar 7h
Netflix US Twitter account hacked t0s.me/2hrFXYG #infosec #cybersecurity #hacking
Retweeted by ISI

Security Ninja @_securityninja yesterday
News in brief: inflight systems can be hacked; LA seeks extradition of cyberattacker;

Facebook

LinkedIn

Subscribe to the DIBBs-ISI Newsletter

Our quarterly newsletter will keep you updated on the datasets that are added each quarter and other changes to the DIBBs-ISI portal.

Join our mailing list

Name:

Email Address:

Re-Enter Email Address:

FIND OUT MORE:

http://www.azsecure-data.org/

http://ai.arizona.edu/research/dibbs