CSAFE All-Hands Meeting 2020

Research on Digital Evidence

Principal Investigators:
Jennifer Newman, Iowa State University
Yong Guan, Iowa State University
Padhraic Smyth, University of California, Irvine

csafe
Center for Statistics and Applications in Forensic Evidence

What is Digital Evidence?

From *Strengthening Forensic Science in the United States: A Path Forward*, The National Acadamies Press, 2009

The analysis of digital evidence deals with…….electronic documents, lists of phone numbers and call logs, records of a device's location at a given time, emails, photographs, and more…..

……In addition to traditional desktop and laptop computers, digital devices that store data of possible value in criminal investigations include cell phones, GPS devices, digital cameras, personal digital assistants (PDAs), ……

Digital Forensics, Statistics, and CSAFE

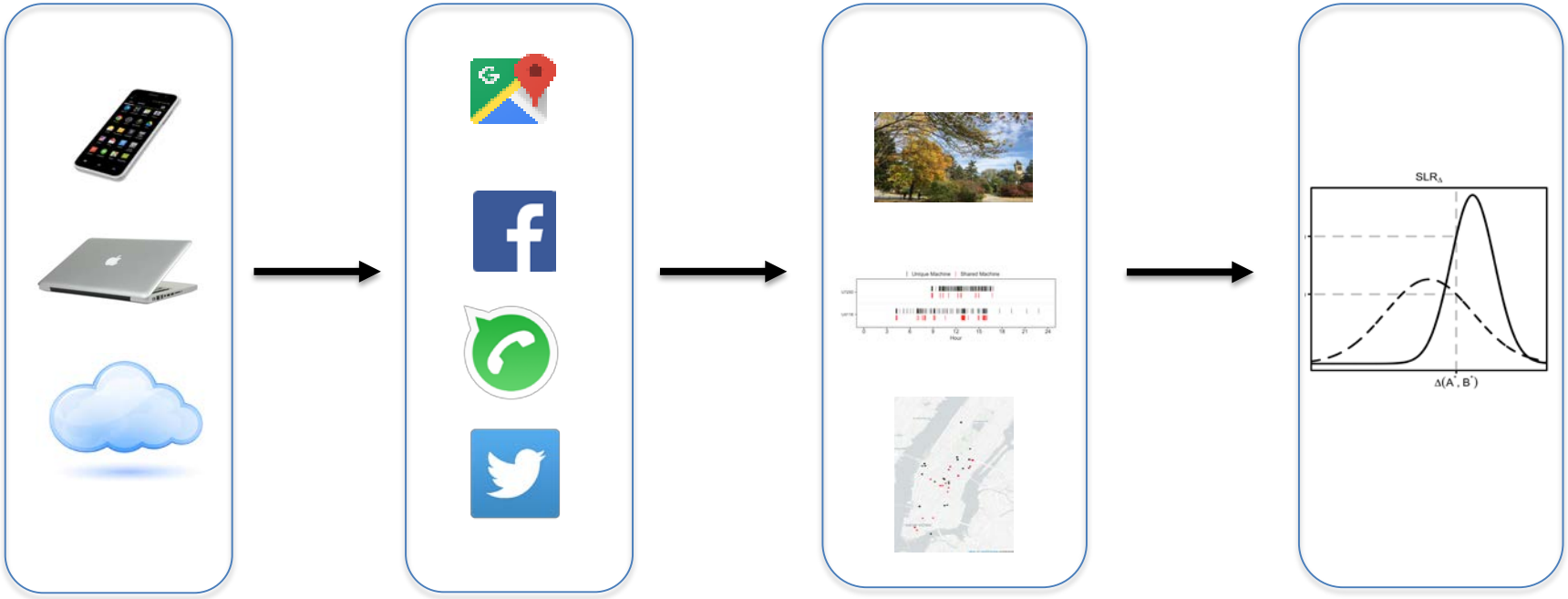
- Volume of digital evidence threatens to overwhelm forensic investigators – widely acknowledged in forensics practice and literature (Roussev, 2016; Vincze, 2016)

- Recommendation from 2019 OSAC R21 Report:
  A Framework for Harmonizing Forensic Science Practices and Digital/Multimedia

  "strengthen scientific foundations of digital/multimedia evidence by developing systematic and coherent methods … to assess the causes and meaning of traces in the context of forensic questions, as well as any associated probabilities."

- August 2014 NIST funding opportunity announcement:
  Center should "*enable the application of probabilistic analysis to two forensic science disciplines: pattern evidence and digital evidence*"

The Process of Digital Forensics



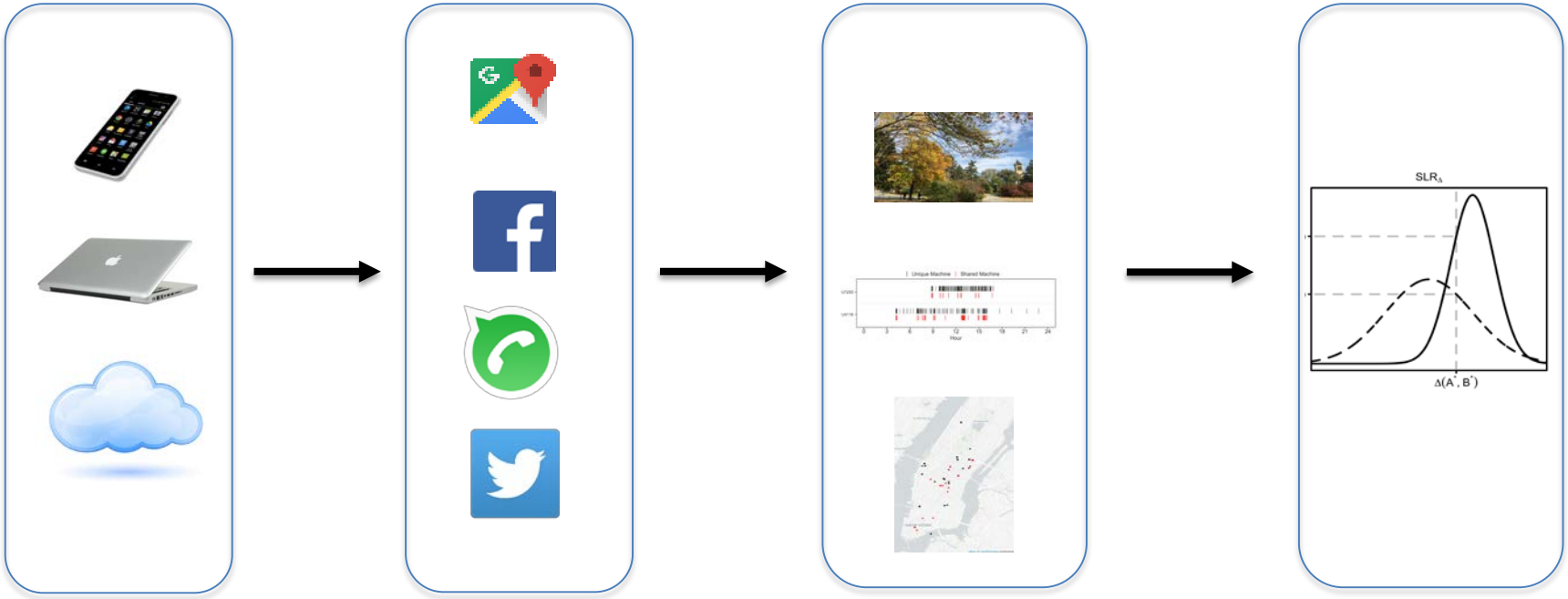
Devices/Accounts Associated with Crimes/Suspects

Extraction of Data and Logged Activities from Software Apps

Analysis of Extracted Data: Images, Text, Events, Geolocation

Quantitative Statistical Inferences

The Process of Digital Forensics



Devices/Accounts Associated with Crimes/Suspects

Extraction of Data and Logged Activities from Software Apps

Mobile App Project
Iowa State

Analysis of Extracted Data: Images, Text, Events, Geolocation

Steganography Project
Iowa State

Quantitative Statistical Inferences

Event Data Project
UC Irvine

csafe
Center for Statistics and Applications in Forensic Evidence

CSAFE 1.0 Projects and Investigators

Research Projects

- **StegoDB: An Image Dataset for Benchmarking Steganalysis Algorithms**
  - Lead PI: Jennifer Newman, Iowa State University

- **Mobile App Evidence Analysis**
  - Lead PI: Yong Guan, Iowa State University

- **Statistical Modeling of User-Generated Event Data**
  - Lead PI: Padhraic Smyth, University of California, Irvine

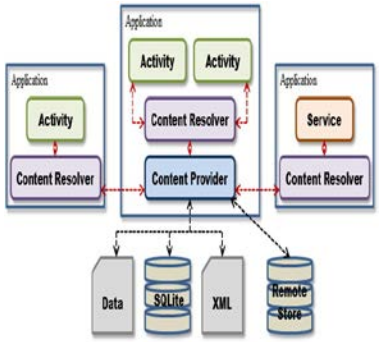
NIST Technical Contact

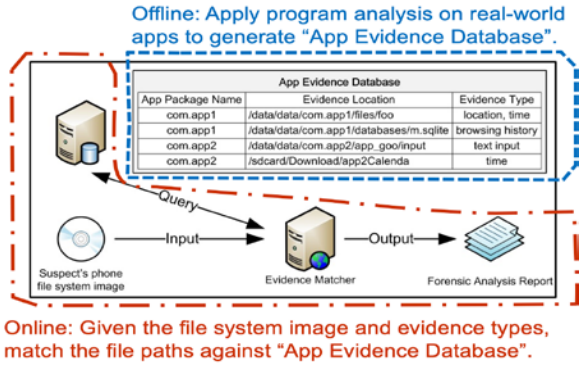
- Barbara Guttman

Project: Mobile App Evidence Analysis Project

PI: Professor Yong Guan, Iowa State University

- **Major Accomplishments**:
  - **Award-winning tools - EviHunter**: Identifying digital evidence from Android devices via static & dynamic Analysis (ACM CCS 2018 and IEEE SADFE 2018)
  - **Android™ App Forensic Artifacts Database**





EviHunter Demo at HFSC, April 17, 2019

- **Impact**
  - **Demo and visit to crime labs**
  - **Invited talks at conferences, webinars for practitioners**

Project: StegoDB: An Image Dataset for Benchmarking Steganalysis Algorithms

PI: Professor Jennifer Newman, Iowa State University

Accomplishments

- Creation of first forensic reference image database for mobile steganography
- > 960,000 images, fully provenanced, publicly available*, copy-right free
- 16 Presentations, 7 publications, 3 PhD students supported, 2 PhD students graduated

Impact

- Using the unique data in **StegoAppDB**
  - Demonstrated no current tools can detect mobile stego images beyond random guess
  - Experimentally verified anecdotal assumption that exposure settings effect the error rates of steganalysis and camera device identification
  - First publication to show that machine learning can detect mobile stego images - data not available until our dataset was created
- Since July 2019, over 5.4 million images downloaded from outside ISU
- Showed a new application of Score-based likelihood ratios (SLR) to camera device identification

csafe
Center for Statistics and Applications in Forensic Evidence

*https://forensicstats.org/stegoappdb/

Project - Statistical Analysis of User-Generated Event Data

PI: Professor Padhraic Smyth, UC Irvine

Collaborators at UC Irvine: Hal Stern (faculty), Chris Galbraith (PhD student)

- **Major Accomplishments:**
  - Developed new statistical methodologies for answering same-source forensic questions from temporal and spatial user-generated event data, e.g., from mobile phones
  - Developed new methods and models for computing likelihood ratios, score-based likelihood ratios, and coincidental match probabilities
  - Conducted systematic evaluation and comparison of these methods on real-world temporal and geolocated event datasets from mobile phones and laptops
    - ROC, TPR/FPR characterization of same/different source pairs from over 1 million mobile device events

- **Impact:**
  - Publications: *Digital Investigation* (2017, 2020), *J. Roy Stat Soc A* (2020)
  - Conference/Workshop Presentations:
    DFRWS (2017, 2020), JSM (2018, 2019), ICFIS (2017), IAFS (2021)
  - Publicly available software: AssocR software package in R, Python code in Github
  - Statistics PhD student joining startup (Obsidian Security) in cybersecurity/forensics

Vision for CSAFE 2.0 Digital Evidence Projects

As our 2.0 projects enter the second phase, we will

- Address issues that are closer to practical considerations of forensic investigators
- Share expertise between PIs

- **Mobile App Artifacts Database (**Guan) can share
  - Expertise from visits with practitioners, mobile apps programming analysis, details on mobile apps artifacts related to event analysis

- **Reference Database StegoAppDB** (Newman) can share/collaborate on
  - Expertise on smartphone photo data, conducting surveys with crime labs

- **Statistical Modeling of Digital Event Data** (Smyth) can share
  - Expertise in applications of statistical models/methods related to digital event data, particularly geolocation data

CSAFE 2.0: Mobile App Evidence Analysis

PI: Professor Yong Guan, Iowa State University

Proposed Activities:

- Test and improve EviHunter
  - Native code and third-party library analysis
  - Packing/unpacking and code obfuscation
  - Survey and analyze evidential artifacts used in real-world cases
- Android App Forensic Artifact Database
- Outreach
  - Workshops and training courses on EviHunter and Android App Forensic Artifact Database

Potential Impact:

- Reduce backlogs at crime labs
- Provide better completeness and accuracy guarantees
- Reducing complexity of mobile device forensic investigation

CSAFE 2.0: StegoAppDB

PI: Professor Jennifer Newman, Iowa State University

- **Proposed Activities**:
  - Acquire several million outdoor images, to provide enough data for deep learning nets
    - Use car, bicycle, (robot?), and in-house Cameraw camera app
  - Create photo-edited images to feed into stego apps, to simulate user actions on phones
  - Use emulators and reverse engineering to provide additional stego apps for embedding
    - Create an in-house detection tool to verify embedding and extraction process
  - Conduct brief survey to assess the status quo of steganography and identify needs of practitioners, create partnerships with forensic practitioners

- **Potential Impact**:
  - larger dataset with photo-edited images gives capability to develop future tools of steganalysis, both academic and practical
    - Simulate how users may create stegos – photo-editing before hiding message
  - Partnership gives greater understanding between two camps on issues of steganography: CSAFE statisticians and forensic practitioners

CSAFE 2.0: Statistical Modeling of Digital Event Data

PI: Professor Padhraic Smyth, UC Irvine

Proposed Activities:

- Develop foundations of quantitative and statistical methods for forensic analysis of event data from mobile devices (geolocation, temporal, metadata)
- Develop new algorithms and software tools for efficient and scalable computation of quantities such as likelihood-ratios for digital event data
- Evaluate and calibrate these algorithms on large real-world data sets
- Make available public-domain testbed datasets and open-access software

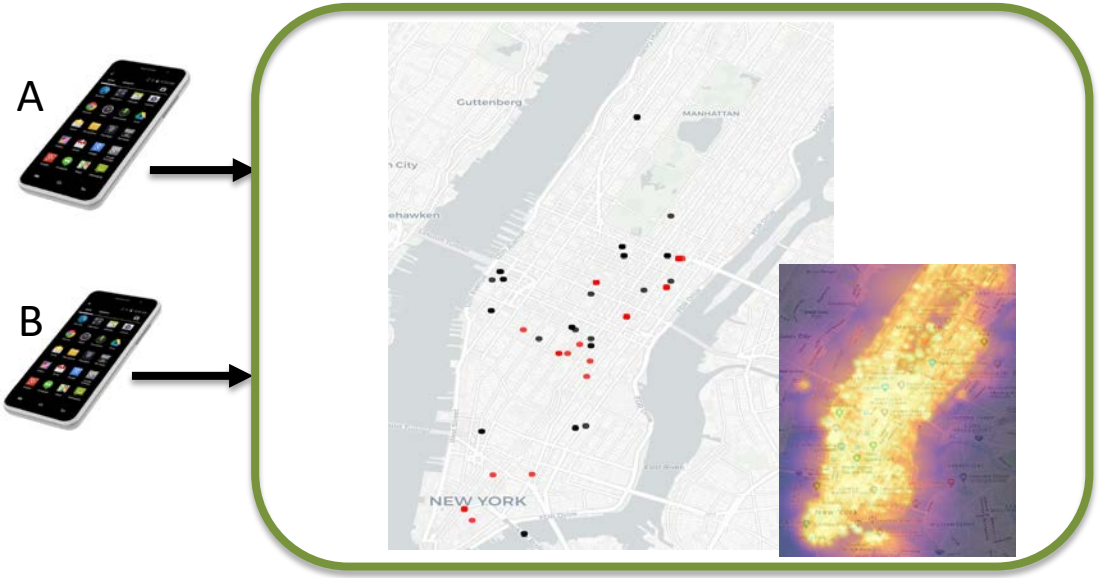
Potential Impact:

- Provide forensic investigators with statistical methodologies for answering forensic questions related to digital event data
- Enable technology transfer to practitioners in digital forensics via software libraries, dataset archives, and tutorial materials

Example: Same or Different Source for Geolocation Data?

A: geolocation data from crime scene
B: geolocation data from suspect

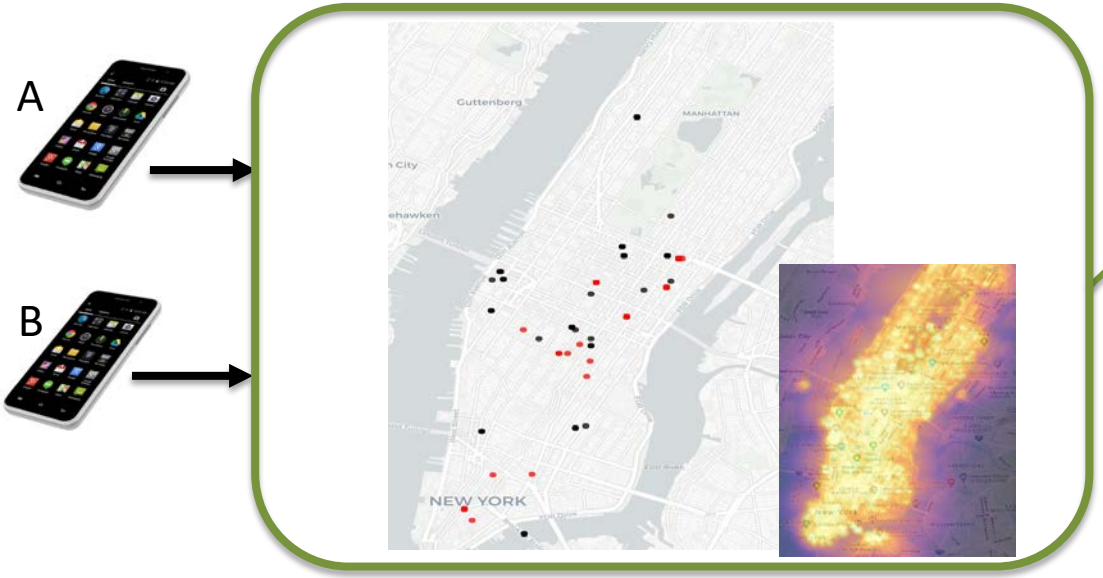
Question of interest:
How likely is it that A and B are from the same source?

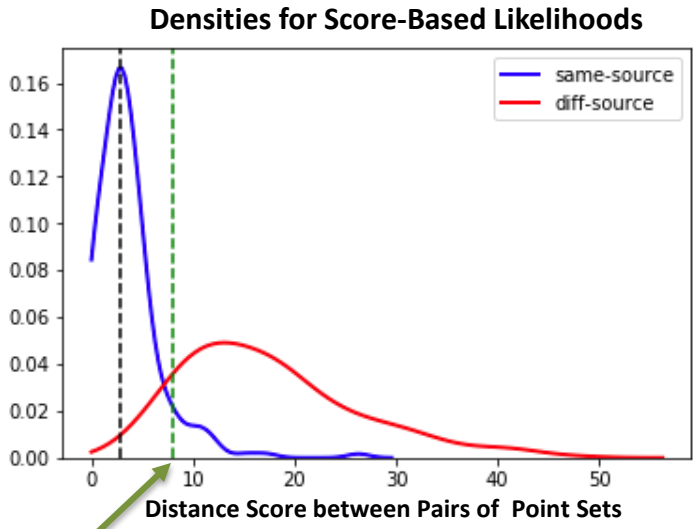


Example: Same or Different Source for Geolocation Data?

A: geolocation data from crime scene
B: geolocation data from suspect

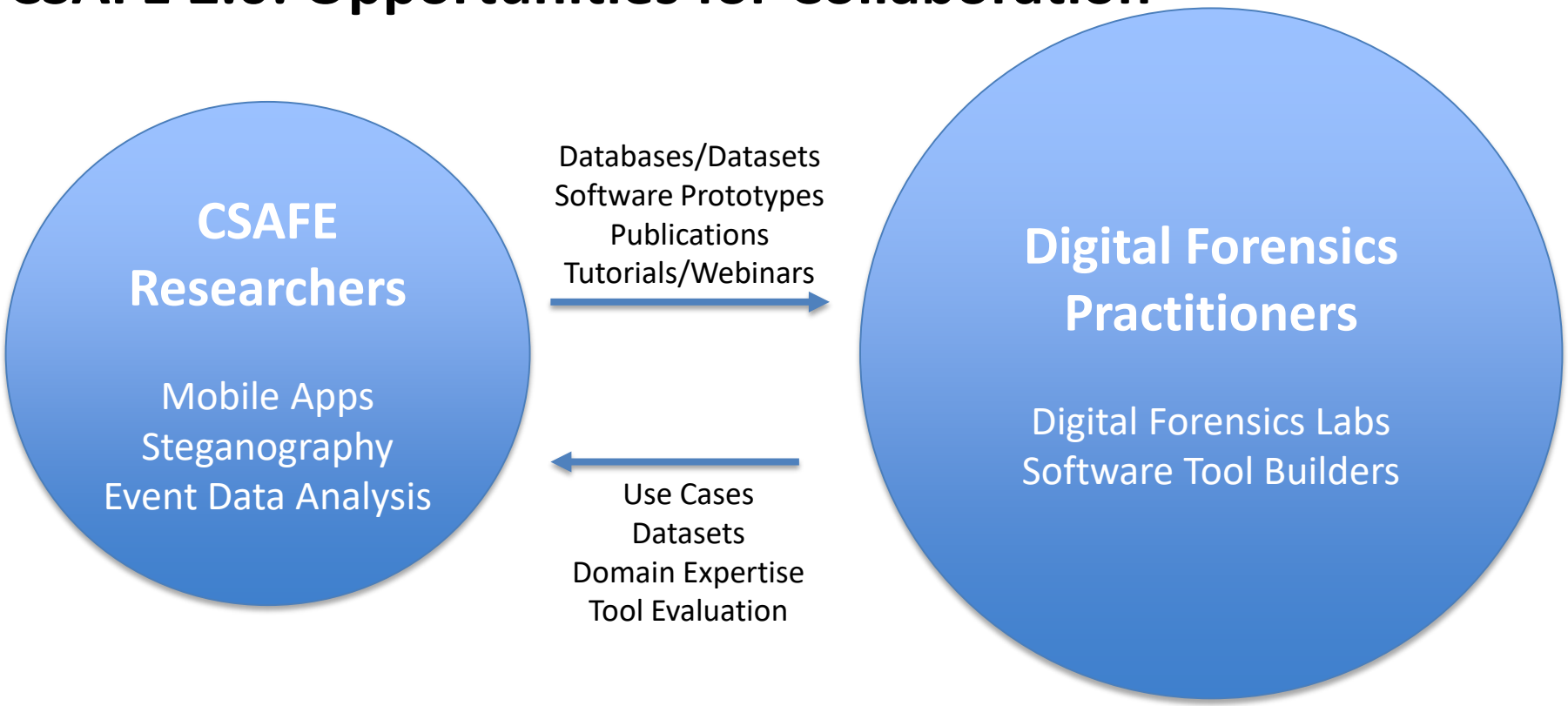
Question of interest:
How likely is it that A and B are from the same source?





See Galbraith, Smyth, Stern, 2020
Statistical methods for forensic analysis of geolocated event data
DFRWS US (upcoming)

CSAFE 2.0: Opportunities for Collaboration



Please contact us if interested:

- **Mobile Apps: Yong Guan, guan@iastate.edu**
- **Steganography: Jennifer Newman, jlnewman@iastate.edu**
- **Event Data: Padhraic Smyth, smyth@ics.uci.edu**

csafe
Center for Statistics and Applications in Forensic Evidence

Questions and Discussion