SecAI: Integrating Cyber Security and Artificial Intelligence/Data Science

+

Diversity and Inclusion

Dr. Bhavani Thuraisingham

The University of Texas at Dallas

September 9, 2020

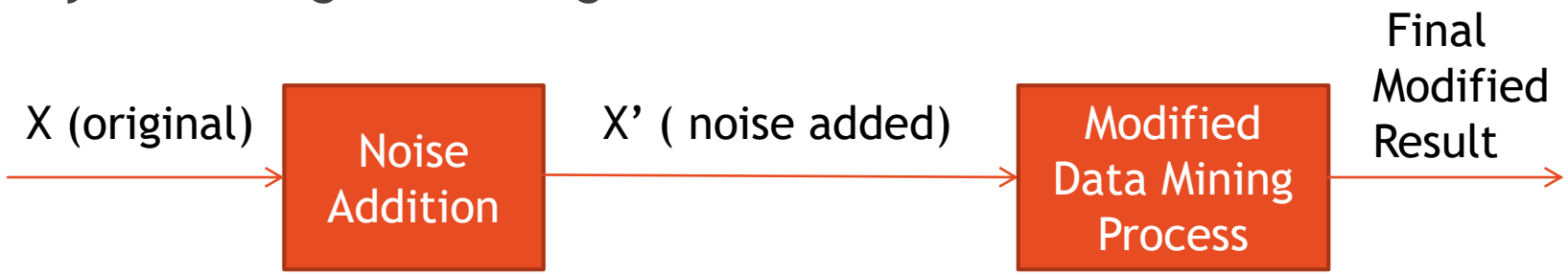
Outline

- Cyber Security and Data Science
  - Motivation
  - Big Data Security and Privacy
  - Privacy Aware Quantified Self
  - Data Science for Cyber Security
  - Adversarial Machine Learning
  - Directions
- Diversity and Inclusion

Acknowledgements: NSF, ARO, AFOSR, NASA, NSA
Colleagues: Profs. Murat Kantarcioglu and Latifur Khan
Project Coordinator: Ms. Rhonda Walls

Motivation: Data Mining, Security and Privacy

- Introduced the idea of Data Mining, Security and Privacy at keynote addresses first at IFIP 11.3 in 1996 and later at PAKDD in 1998 and subsequently published a landmark paper* while at NSF in 2002 that spawned a new area of research.
- Privacy-Preserving Data Mining



- Our research with PhD student Dr. Li Liu focused on Privacy-Preserving Decision Trees and the Perturbation Method** between 2005 - 2008.
- Data Mining, Security and Privacy is exacerbated with Big Data and Data Science; hosted an NSF Workshop on Big Data Security and Privacy in September 2014 and presented the results to the Interagency Working Group in February 2015.

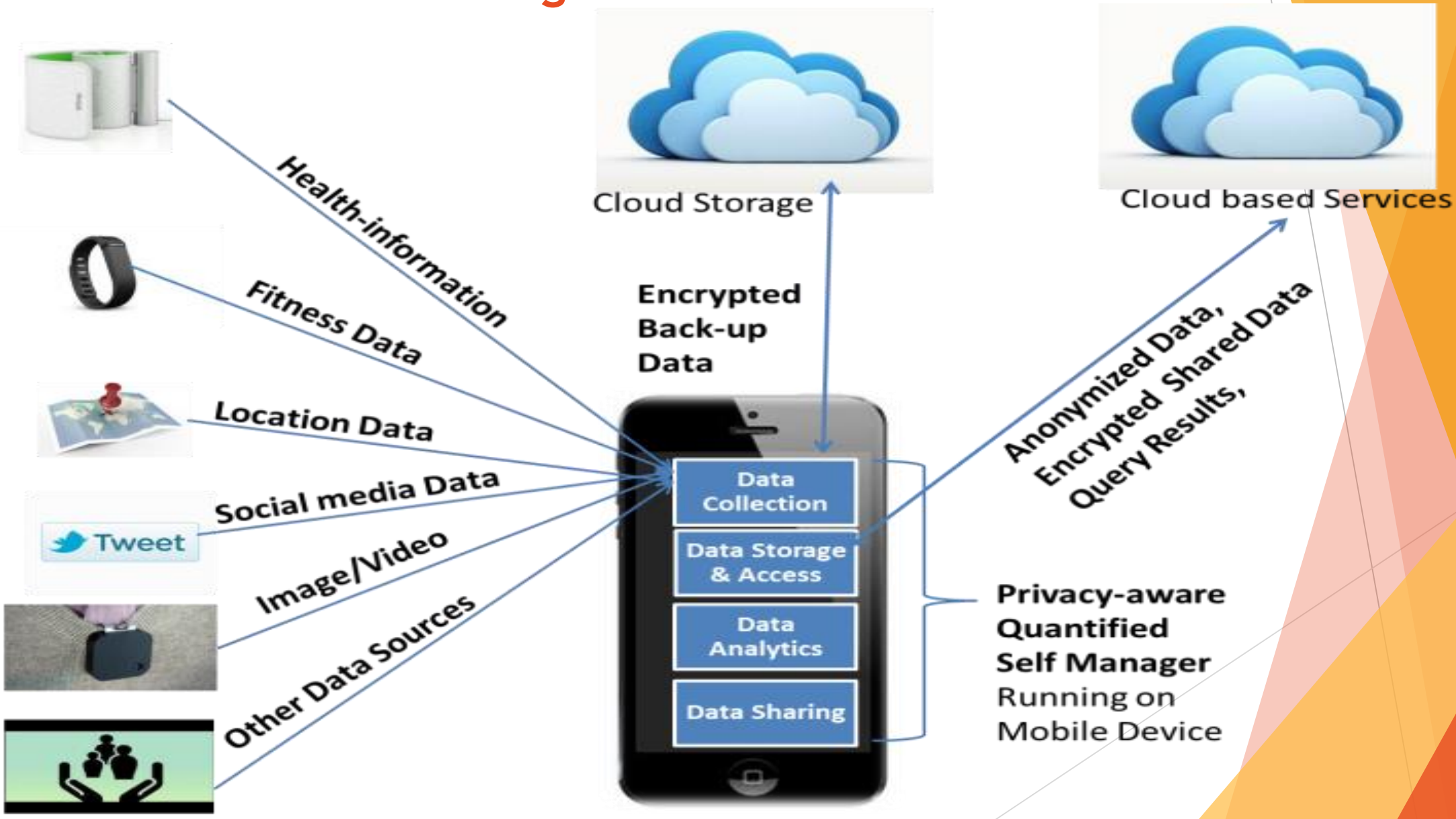
* Bhavani M. Thuraisingham: Data Mining, National Security, Privacy and Civil Liberties. SIGKDD Explorations 4(2): 1-5 (2002)

**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): 5-21 (2008)

Big Data Management, Security and Privacy

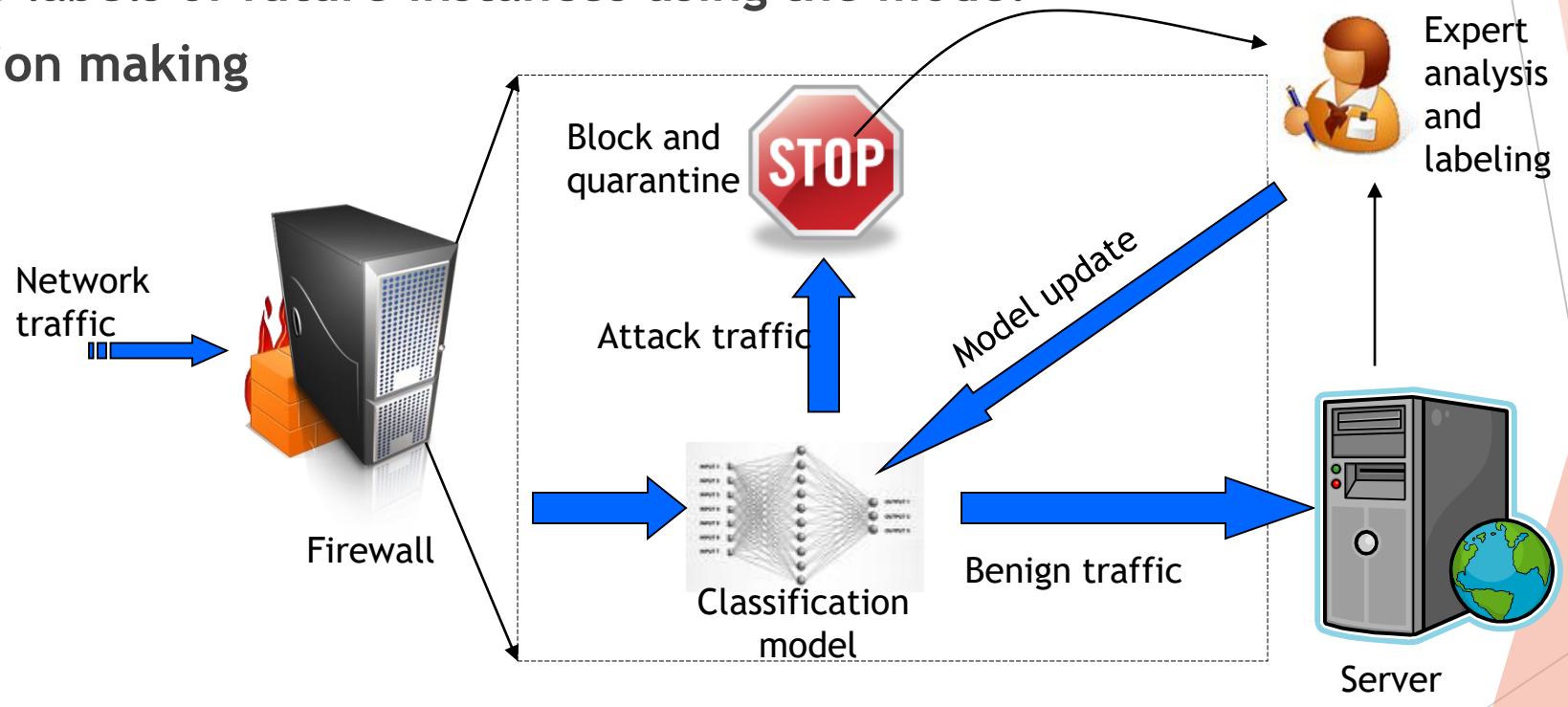
- Due to technological advances and novel applications, it is possible to capture, process, analyze large amounts of data for security tasks.
- Such tasks include user authentication, access control, anomaly detection, user monitoring, and protection from insider threat.
- By analyzing and integrating data collected on the Web, one can identify connections and relationships among individuals that may in turn help with homeland protection, disease outbreaks, ….
- Collected data, even if anonymized by removing identifiers, when linked with other data, may lead to re-identifying the individuals.
- Security tasks such as authentication and access control may require detailed information about users (e.g., multi-factor authentication).
- This information, if misused or stolen, can lead to privacy breaches.
- Directions include Securing the Data – Access Control Models, Privacy Enhanced Techniques, Big Data Analytics for Cyber Security.

Privacy-Aware Policy-based Quantified Self: Data Management Framework



Data Science for Cyber Security Applications
Big Data Stream Classification* (with Prof. Latifur Khan)

- Uses past data to build classification model
- Predicts the labels of future instances using the model
- Helps decision making



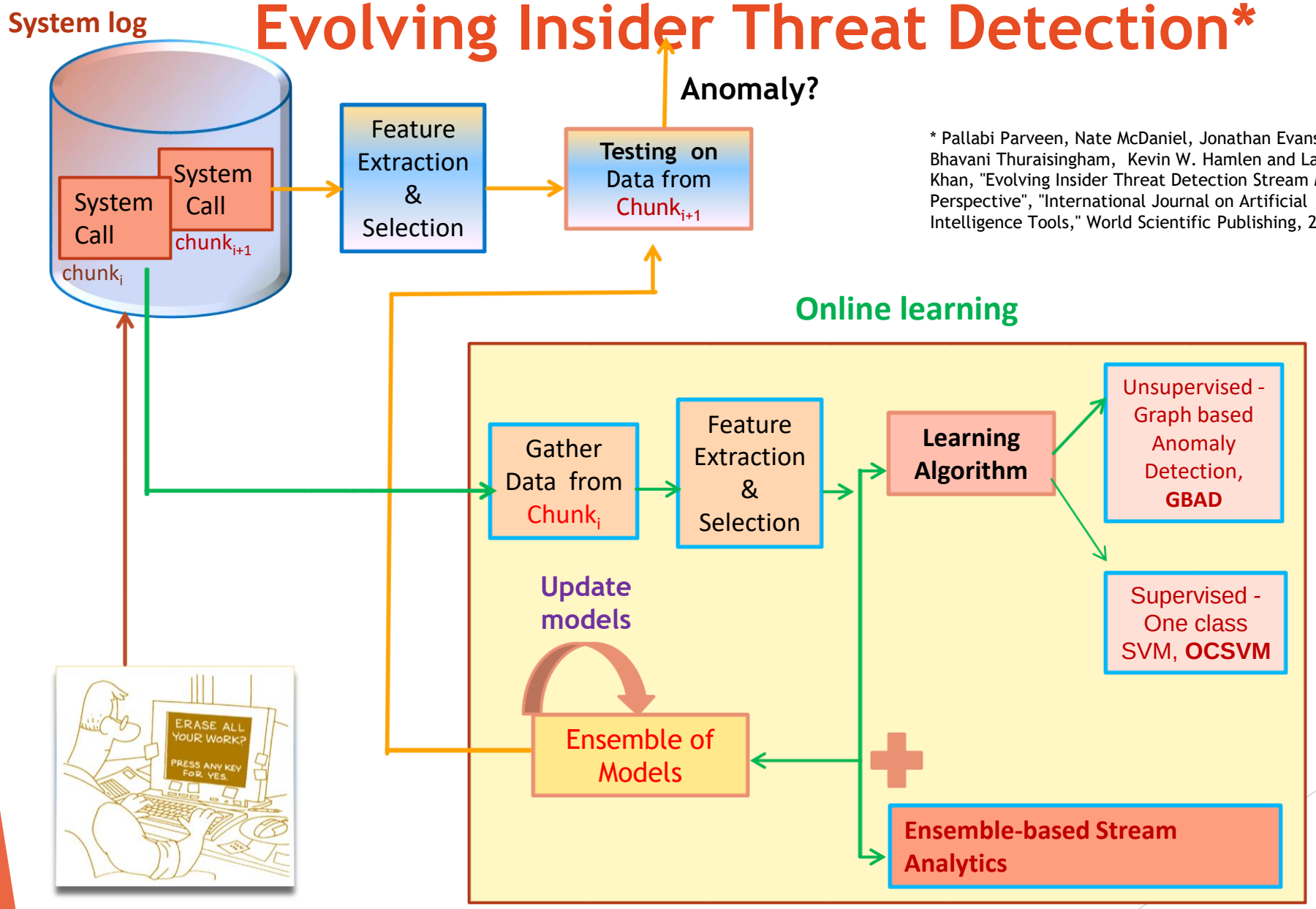
Big Data Streams:

- are a continuous flow of data
- are very common in our connected digital world
- have massive amounts of data

*Mehedy Masud, Latifur Khan and Bhavani Thuraisingham, Data Mining Tools for Malware Detection, CRC Press, 2011.

Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, Bhavani M. Thuraisingham: Classification and Novel Class Detection in Concept-Drifting Data Streams under Time Constraints. IEEE Trans. Knowl. Data Eng. 23(6): 859-874 (2011)

Application: Architecture for Evolving Insider Threat Detection*



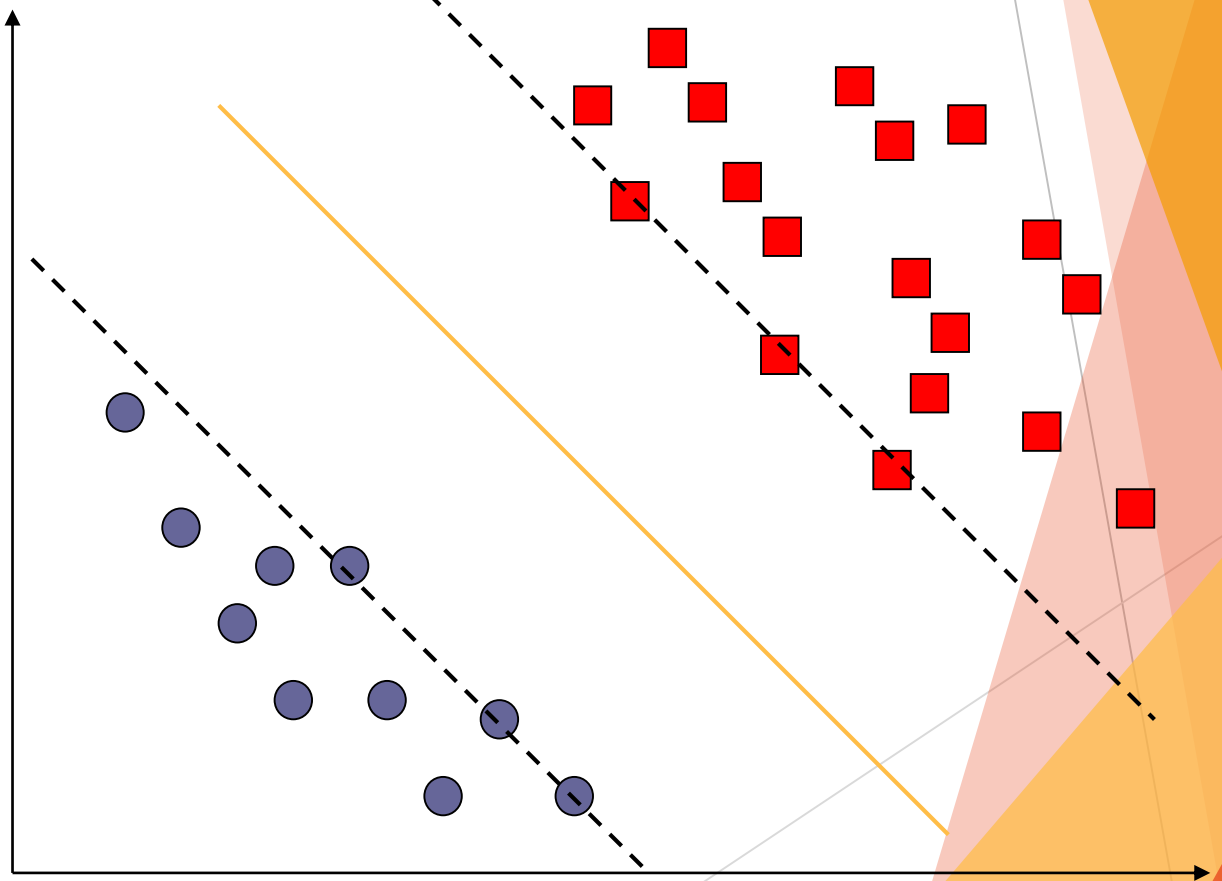
* Pallabi Parveen, Nate McDaniel, Jonathan Evans, Bhavani Thuraisingham, Kevin W. Hamlen and Latifur Khan, "Evolving Insider Threat Detection Stream Mining Perspective", "International Journal on Artificial Intelligence Tools," World Scientific Publishing, 2013

Securing Data Science/ML/AI
Adversarial Machine Learning: The Problem
(with Prof. Murat Kantarcioglu)

- Adversary modifies data to defeat learning algorithms

Understanding Adversarial Learning

It is not concept drift
It is not online learning
Adversary adapts to avoid being detected
    During training time (i.e., data poisoning)
    During test time (i.e., modifying features when data mining is deployed)
There is game between the data miner and the adversary



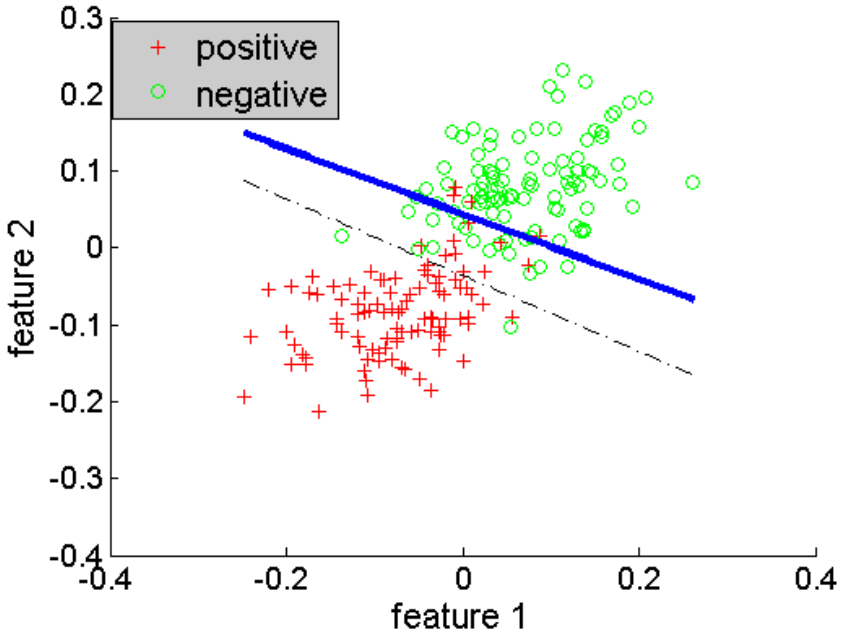
Adversarial Machine Learning
AD-SVM* Example:

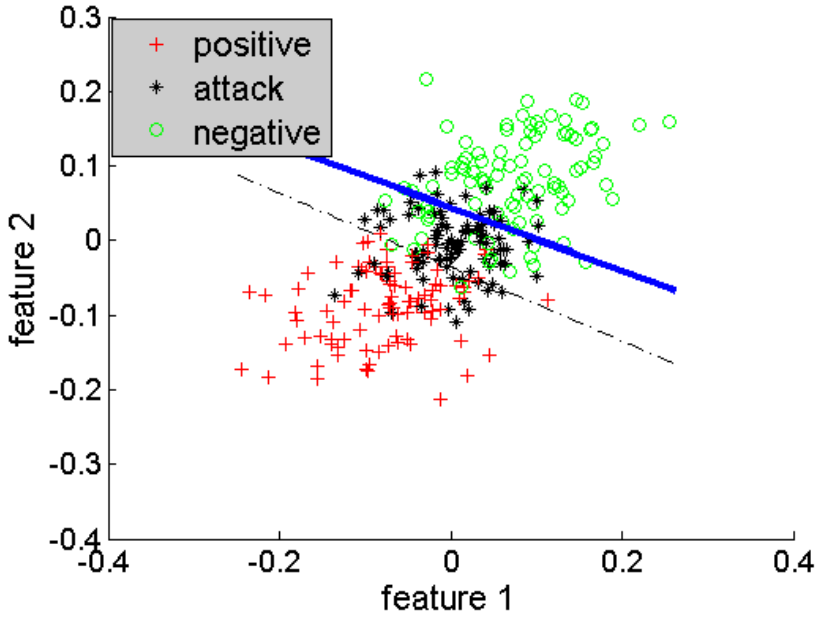
Threat Model Example

Test time/ Deployment Time Attacks

Attacker modifies x to x'

- Modify packet length by adding dummy bytes
- Add good word to spam e-mail
- Add noise to an image



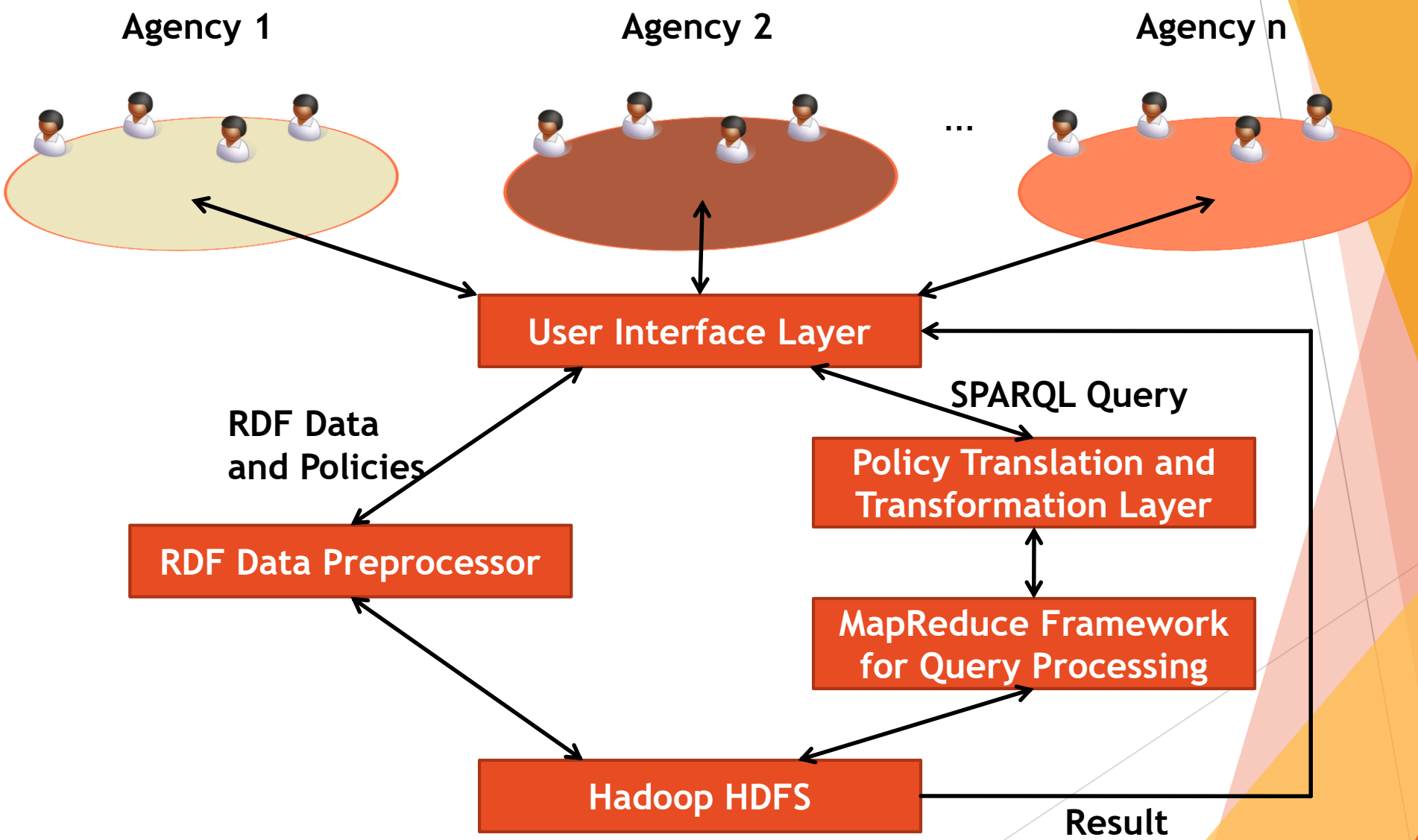


black dashed line is the standard SVM classification boundary, and the blue line is the Adversarial SVM (ADV-SVM) classification boundary

*Y. Zhou, M. Kantarcioglu, B. Thuraisingham, and B. Xi, Adversarial support vector machine learning, ACM SIGKDD '12

One of the handful of early efforts on Adversarial Machine Learning for Cyber Security

Cloud-based Assured Information Sharing



Directions

- Developments in Integrating Cyber Security and Artificial Intelligence (Data Science, ML) are exploding.
- Applications of AI for Security (Insider Threat Detection) and Security for AI (Adversarial Machine Learning)
- Also, developments in Privacy Aware Healthcare and managing one's life activities: Privacy Aware Quantified Self
- At the same time, with the development of the Internet of Things including Internet of Transportation Systems and Infrastructures, security and privacy solutions are being developed, mainly physics-based solutions.
- As driverless cars become a reality, more and more of the Autonomous Vehicles (AVs) will use AI/ML techniques.
- How can we develop AI solutions for detecting malware in AV systems and how can Adversarial ML work in AV systems?
- Can we develop a Privacy-Aware Policy based Data Management Framework for Internet of Transportation and Infrastructures?
- Is it possible to perform trustworthy analytics efficiently when so much of data is coming from all directions?
- Numerous opportunities for substantial research and development!

Diversity and Inclusion

- The concept of **multicultural and diversity management** encompasses acceptance and respect, recognition and valuing of individual differences. Diversity is defined as differences between people, that can include dimensions of race, ethnicity, gender, sexual orientation, socioeconomic status, age, physical abilities, religious beliefs, political beliefs, or other ideologies. Multiculturalism refers to the existence of linguistically, culturally and ethnically diverse segments in an organisation.

- https://en.wikipedia.org/wiki/Multicultural_and_diversity_management

What is Unique About Cyber Security and AI/Data Science

- **They are two of the more lucrative fields in Computer Science with high income potential**
- **Cyber Security and Data Science/AI researchers and developers are in great demand in Government, Academia, and Industry**
- **There is substantial funding with many Agencies both in Research and Education and this trend is expected to grow**
- **Women still make up less than 20% in many Computer Science fields (e.g., Cyber Security around 10% and Data Science/AI around 25%; Percentage of Under-represented Communities is even lower.**
- **Cyber Security and AI help humanity and are Intellectually Challenging fields.**

10 Reasons: Why a Career in Cyber Security and/or Data Science/AI for Female and Under-Represented Communities

- **Given the opportunity Women/Under-Represented communities can excel in any Computing field and Cyber Security and Data Science/AI are very exciting fields with so many innovations and developments happening so rapidly.**
- **It can be integrated with many areas – arts, humanities, natural science, social science, engineering, business, medicine and law with the need for IoT and related technologies.**
- **There are many options from research and academia to product development to start-ups with substantial funding.**
- **Millennial Women/Under-Represented communities and beyond have the flexibility and freedom to choose careers and have Female/Under-Represented role models in the field that us baby boomers never had.**
- **In many research areas you can work from home most days making it ideal especially for Women to have a family and career.**

10 Reasons: Why a Career in Cyber Security/Data Science Career for Female and Under-Represented Communities?

- **Many Cyber Security and Data Science/AI jobs cannot be overtaken by robots – we need researchers and developers to collect and analyze massive amounts of data from the IoT systems.**
- **Cyber Security / Data Science are highly paid fields with numerous job opportunities; why not Female/Under-Represented communities take advantage of these benefits?**
- **Female/Under-Represented communities can be financially independent with a career in Cyber Security/Data Science; financial independence means self-respect, less stress and confidence; Having financial independence is a must for everyone especially for a woman.**
- **Computing systems are everywhere from North to South and East to West – therefore these systems can be hacked. You can make the world a better and safer place with Cyber Security including addressing the major problem of "Violence against Women and Children" with Technology.**

Advice to All Female and Under-Represented Communities

- **Do not be undermined by others – that is when they say "Women and/or Under-Represented communities are not good in STEM"**
- **Never give up when others (even Women) put you down**
- **Work hard, set goals, but be flexible as life does not turn out the way we want it to, especially with marriage and children**
- **Love your work, but also enjoy life**
- **Be a role model for the younger Women and those from the Under-Represented communities and learn from the successful Women**
- **Even though you may get into administration which is very important to support Women, keep yourself technical and continue to do top quality research and development**

Some Suggestions for Senior Technologists from the Female and Under-Represented Communities

- **Support Female/Under-Presented communities on issues of importance – e.g., tenure related matters, promotions, corporate boards.**
- **Connect technologists from the Female and Under-Represented communities with senior personnel both men and Women.**
- **Mentor Female/Under-Represented communities and give advice on career development.**
- **Promote Female/Under-Represented communities for program committees, program chairs and awards/Fellows (e.g., ACM and IEEE Fellow), venture funding.**
- **Promote Female/Under-Represented communities to take on administrative positions – very important so that they can mentor junior technologists.**
- **Introduce Computer Science Education in Elementary, Junior and Senior High Schools.**
- **We need more role models!**

Learn from our Female and Under-Represented Communities role models - Sample (even if not all of them are in STEM)

- Prof. Alan Turing (who can forget him - he gave us computer science)
- Countess Ada Lovelace (she gave us programming)
- Admiral Grace Hopper (we remember her through GHC)
- Actress Hedy Lamarr (who can forget the beautiful and brilliant Ms. Lamar who gave us wireless)
- Self-made billionaire Oprah Winfrey (A role model for all Women)
- Tennis Player Serena Williams (Greatest of all times)
- And we have our own STEM role models at UTD - Provost Dr. Inga Musselman and Dean of Engineering and Computer Science Dr. Stephanie Adams as well as Dean of Graduate Education Dr. Juan Gonzalez.
- .....

Contact

- Prof. Bhavani Thuraisingham, bhavani.thuraisingham@utdallas.edu
- Prof. Murat Kantarcioglu, muratk@utdallas.edu
- Prof. Latifur Khan, latifur.khan@utdallas.edu
- Ms. Rhonda Walls, rhonda.walls@utdallas.edu