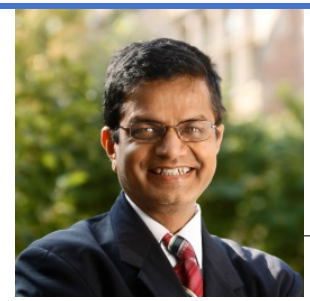
Our Research on New AI, and Blockchain Techniques for Network Security



Raj Jain

Washington University in Saint Louis

Saint Louis, MO 63130

Jain@wustl.edu

A talk in "CSE 591: Introduction to Graduate Studies in CSE"

September 30, 2022

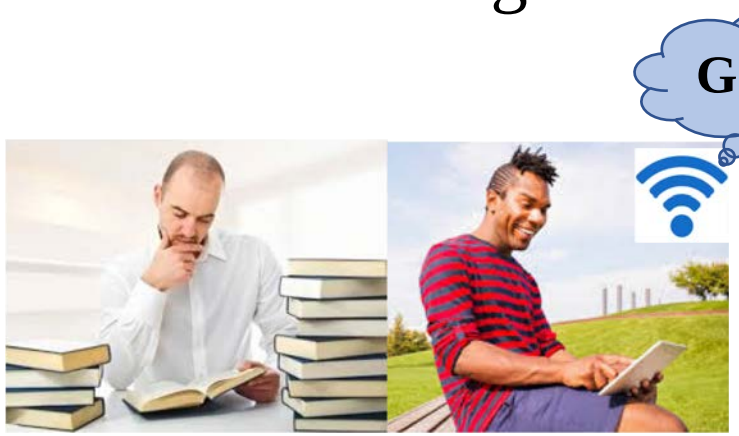
These slides and a video recording of this talk are at:

http://www.cse.wustl.edu/~jain/talks/cs59122.htm

1. Trends in:
   - AI
   - Blockchain
   - Cybersecurity
2. Our Research in these areas
3. Key distinctions of our research

Past: Smart Things

- IoT = Internet of Things = Connected Things
- Things=Anything (other than computers)
- Google made worldwide information retrieval instantaneous
- Instant knowledge ⇒ Smart



Not-Smart Smart

Smart Watch

Smart Health

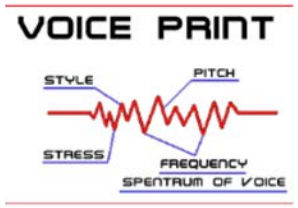
Smart Home

Smart Cities

Present: Intelligent Things

- Recently, AI became a reality
  (the concept of AI has been around since Turing's time but was limited)
- Trend: Smart ⇒ Intelligent (Like humans with five senses)
- Devices that can figure out what they touch, see, or hear
  (Smell and taste are still in research)
  Simple pattern recognition ⇒ intelligent touch/visual/sound recognition

Touch ID



Voice Print

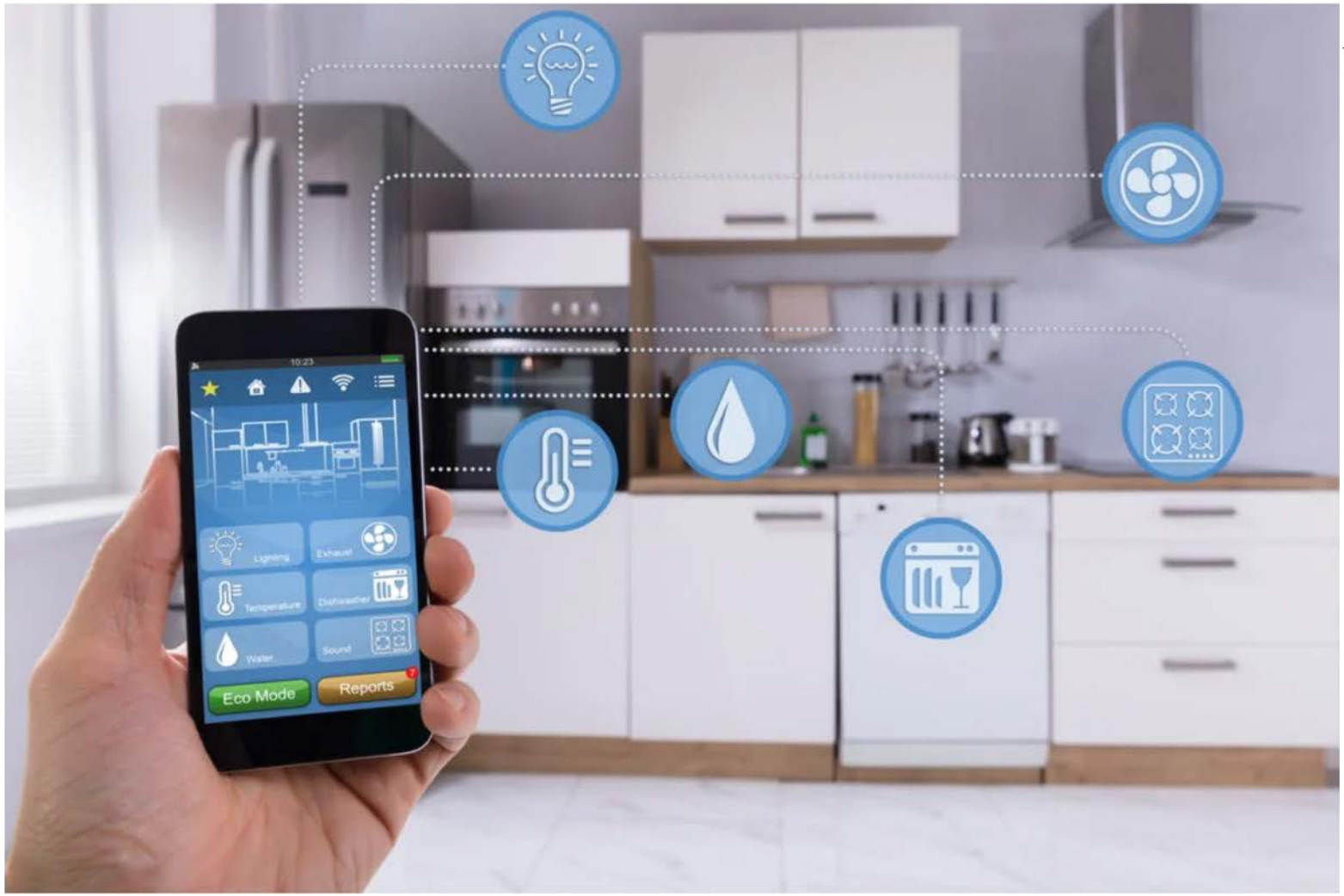
Facial Recognition

Amazon Alexa

Self-driving Car

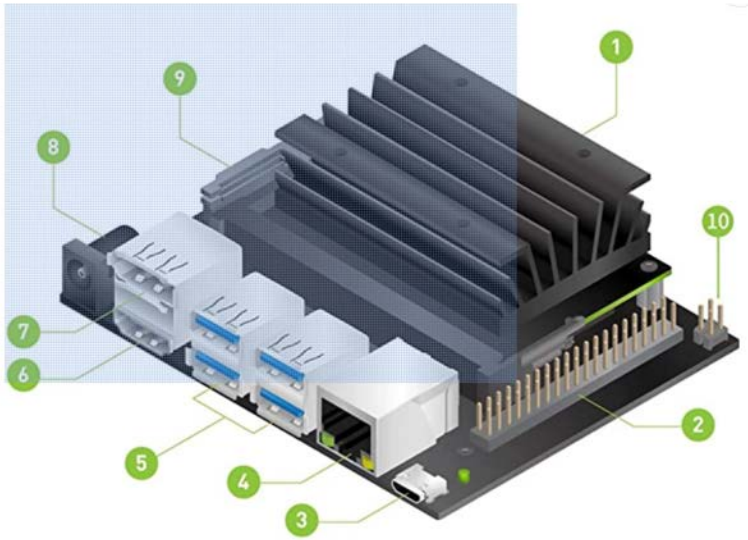
AI is everywhere

- Coffee Machines
- Vacuums
- Manufacturing Robots
- Self-Driving cars
- Self-Driving Networks
- Plane Auto-Pilot



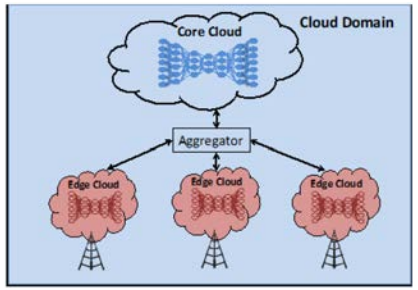
Coming: Edge Intelligence

- AI requires a significant amount of computation ⇒ Intelligence is in the cloud
- Only simplified recognition can be done in the devices
- Moore's law and miniaturization of electronics are bringing cloud-like capabilities to the edge
- Edge is moving close to the devices ⇒ Self-Intelligent devices
  AI in end systems, gateways, servers, and clouds
- AI ASICs (Google Coral, Nvidia Jetson nano) are bringing AI to devices ≈ $200



Source: https://www.amazon.com//dp/B09T37PPRF

Also coming:



1. **Distributed/Federated AI**
2. **Transformers**: DNN models used in language translations
3. **Composite AI**: Using both existing techniques and machine learning
4. **Human-Centered AI**: Augmented intelligence with humans in the loop
5. **Responsible AI**: Fair, unbiased, explainable, regulation compliant
6. **Generative AI**: Generate new ideas, methods, knowledge
7. **AI TRISM** – AI Trust, Risk, Security Management (Explainability, adversarial attack resistance, data protection, anomaly detection, etc.)
8. **Artificial General Intelligence**: Human-like - reasoning, emotions, bias, …grow up more than 18 months old child

AI and IoT Research Funding

- Artificial Intelligence can be used to generate *Artificial Knowledge* fast
  ⇒ Lots of papers and research funding
- In 2019, news media reported China is ahead of the US in AI research
  ⇒ National AI Initiative Act 2020
  ⇒ Billions of dollars on AI. $20M each for several AI institutes
- European Union's Framework Program 7 (FP7) was the first to fund IoT research
- We discovered this during our research on "Next Generation Networks
  ⇒ IoT research since 2010.

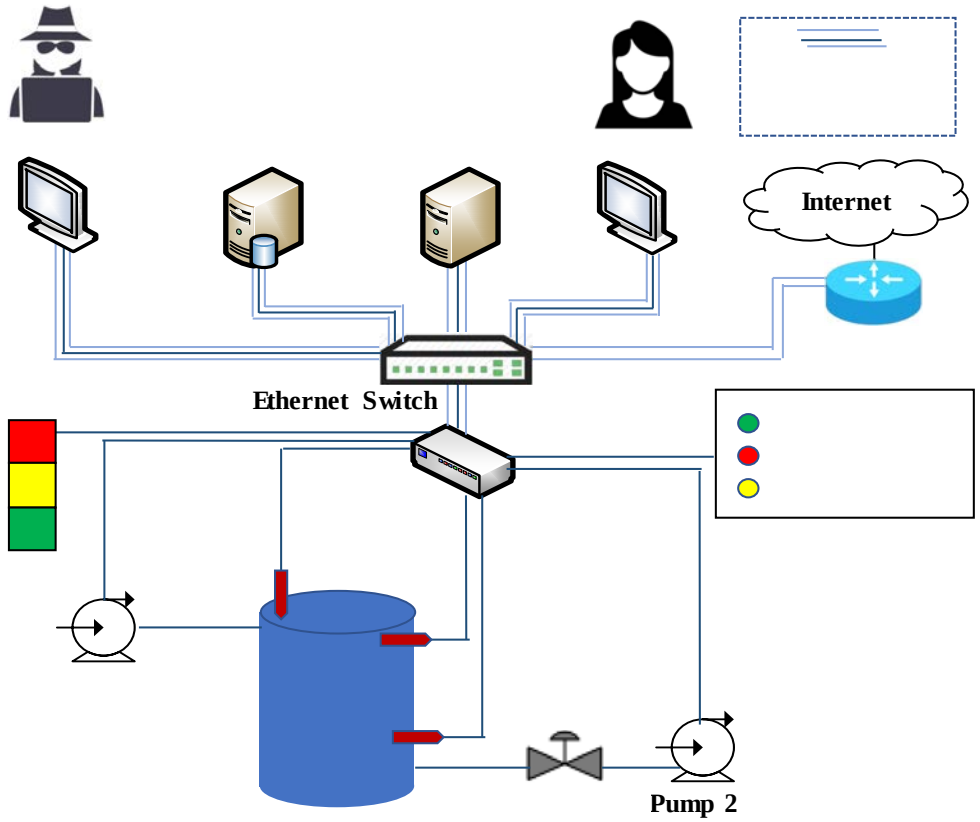
Artificial Intelligence can be used to generate Artificial Knowledge fast

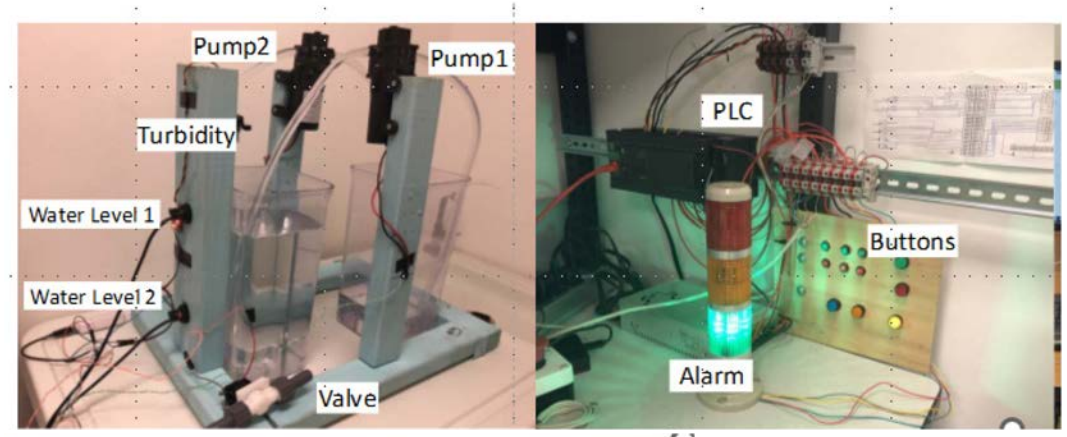
Ref: Subharthi Paul, Jianli Pan, Raj Jain, "**Architectures for the Future Networks and the Next Generation Internet: A Survey**," Computer Communications, UK, Volume 34, Issue 1, 15 January 2011, pp. 2-42, http://www.cse.wustl.edu/~jain/papers/i3survey.htm [299 citations]

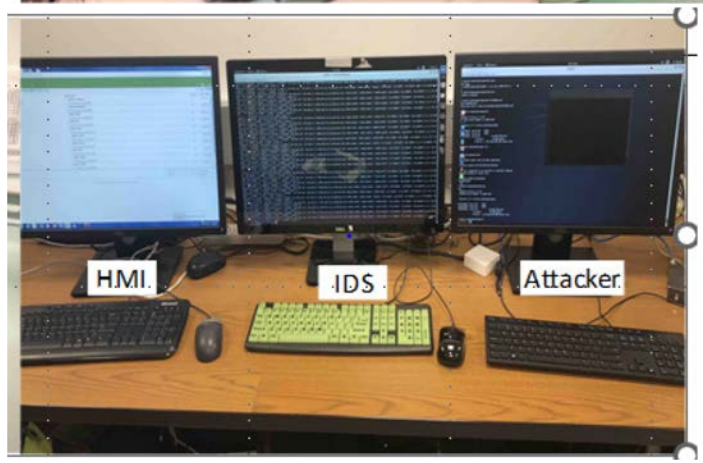
AI-Based Security of IoT: Our Research

- Security research since 2009
- AI research since 2017
- Security of Industrial Internet of Things (IIoT)
- Security of Internet of Medical Things (IoMT)
- Security using blockchains
- 24+ papers
- AI = Pattern recognition, probabilistic reasoning, machine learning, deep learning, …
- Everything we say applies to all of these variations.

Industrial Control Systems Security Using AI

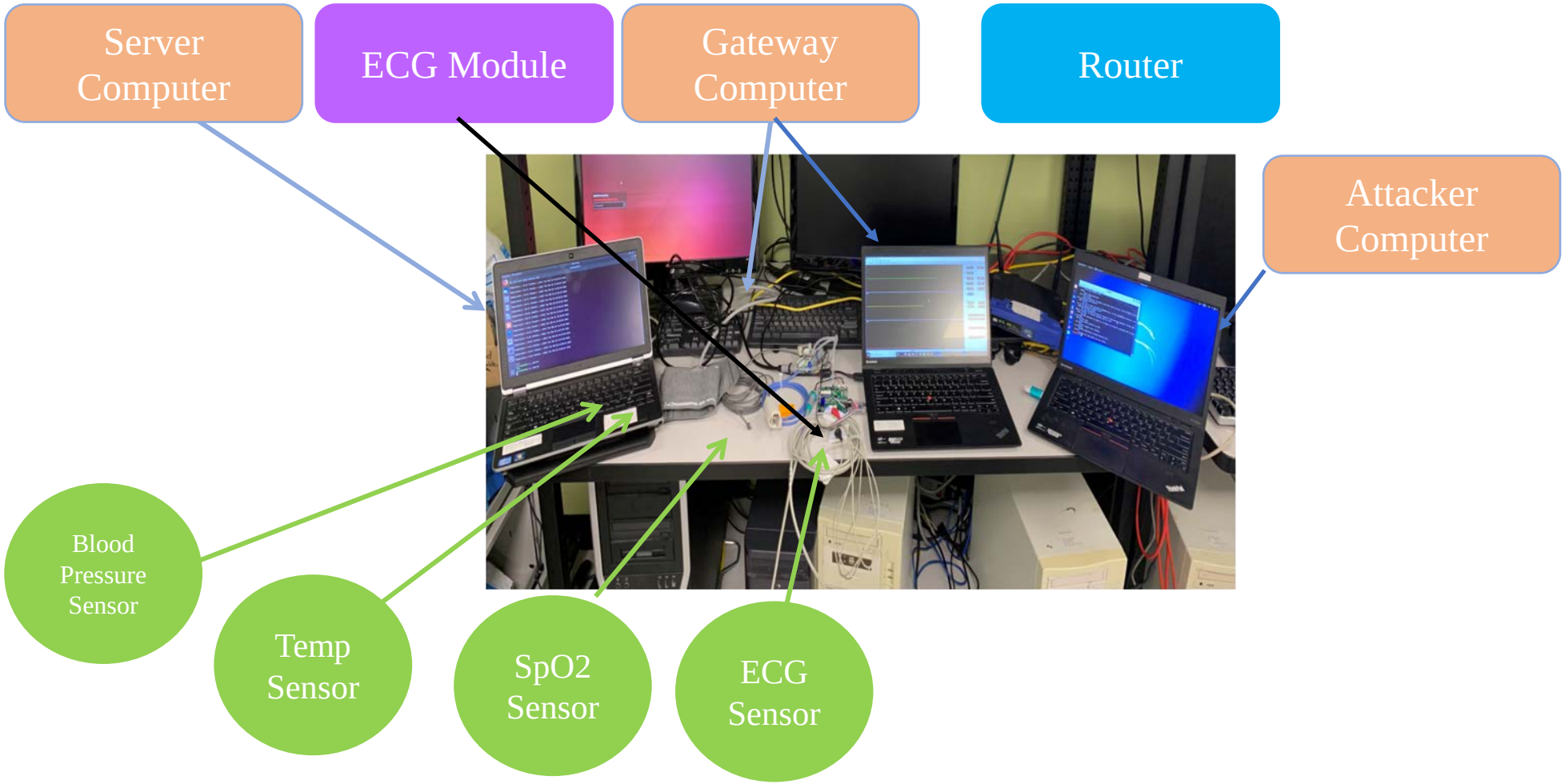






WUSTL-IIOT-2021 Dataset for IIoT Cybersecurity Research, http://www.cse.wustl.edu/~jain/iiot2/index.html

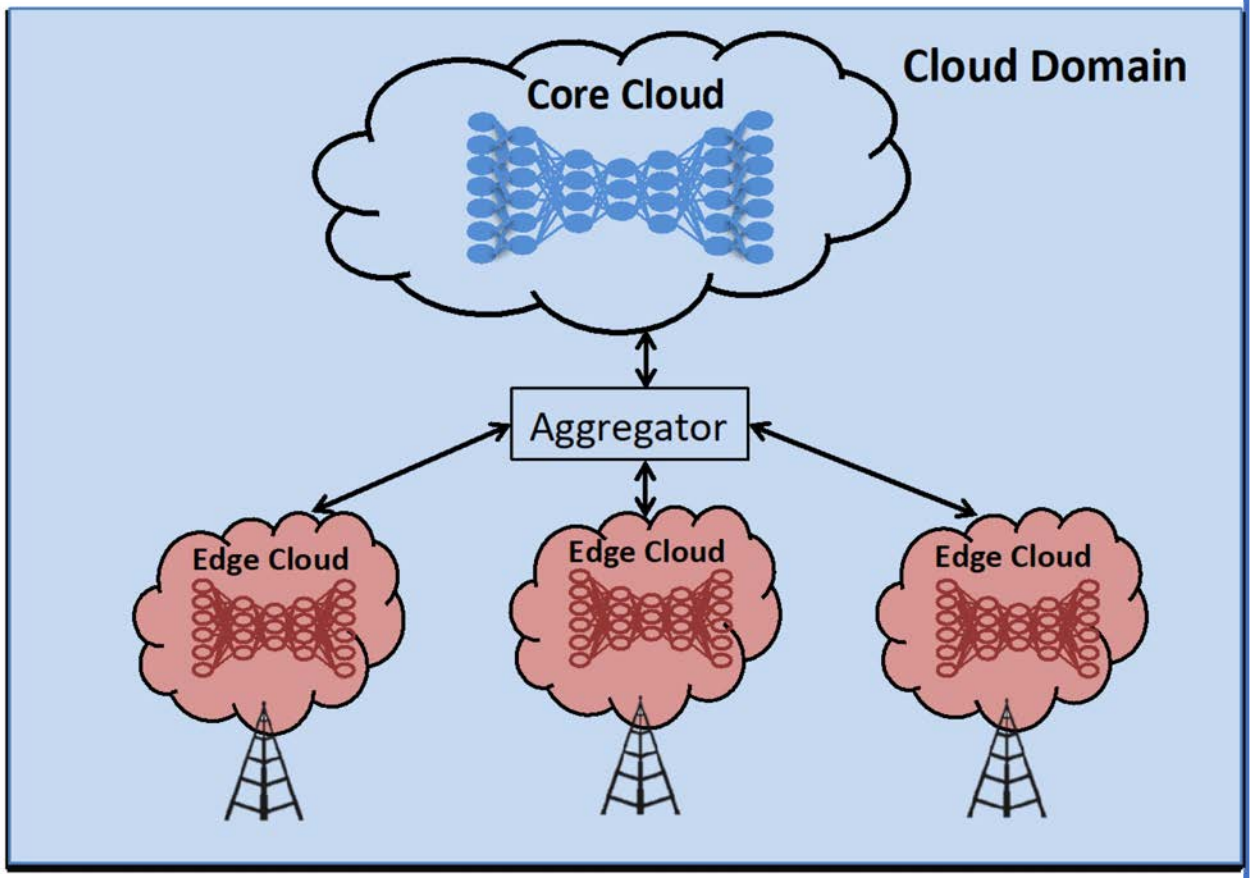
Internet of Medical Things Security Using AI



WUSTL EHMS 2020 Dataset for Internet of Medical Things (IoMT) Cybersecurity Research, http://www.cse.wustl.edu/~jain/ehms/index.html

Edge AI: Hierarchical Deep Learning

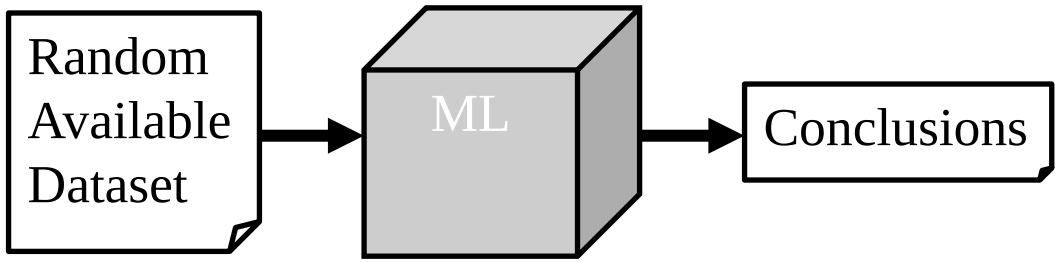
- No need to send data to the core cloud
- Edge clouds send a preliminary model to the core
- Also known as "Federated Learning"



Ref: L. Gupta, T. Salman, A. Ghubaish, D. Unal, A. K. Al-Ali, R. Jain, "**Cybersecurity of multi-cloud healthcare systems: A hierarchical deep learning approach**," Applied Soft Computing (2022), 5 January 2022, http://www.cse.wustl.edu/~jain/papers/muse.htm

2. Random Datasets

- Real data is usually private. Not published.
- Published data is either old or too generic.
- KDD, a commonly used dataset in intrusion studies, is a simulated dataset from 1999.

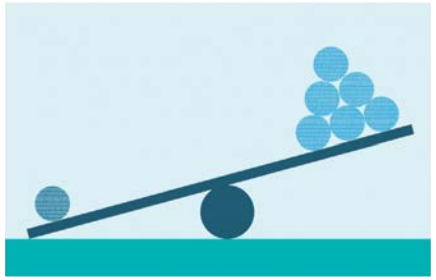


Garbage-In, Garbage-Out

Ref: KDD Cup 1999 Data, October 28, 1999, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

3. Imbalance of Security Data

- AI started with image analysis but needs to be extended for security
- Security data is very different from image data
  - Most security datasets are not representative of the real world.
  - In most papers, 10-15% of the packets are attack packets
- In real-world, 1 in several billion packets is an attack packet
  - Mis-classify the attack packet ⇒ 99.9999% accuracy
- **Extreme Data imbalance** is a critical issue in security



1% attack

Ref: Maede Zolanvari, Marcio A. Teixeira, Raj Jain, "**Effect of Imbalanced Datasets on Security of Industrial IoT Using Machine Learning**," 2018 IEEE International Conference on Intelligence and Security Informatics (ISI), Miami FL, Nov. 9 - 11, 2018, 6 pp., http://www.cse.wustl.edu/~jain/papers/imb_isi.htm

4. Wrong Metrics

- In Image analysis:
  Cost of predicting "0" when it is "1" = Cost of predicting "1" when it is "0."
  ⇒ Cost of errors is symmetric ⇒ Almost all metrics are symmetric.
- In Cyber Security:
  - Cost of missing an attack = 10^6 × Cost of false attack prediction
  - Washington Post (5/30/22): 5 missiles hit Iraqi base hosting US troops
  - Would you live at the base protected with 90% accuracy?
- Need new metric to find the best algorithm ⇒ Use **Safety Score**

Ref: Tara Salman, Ali Ghubaish, Devrim Unal, Raj Jain, "**Safety Score as an Evaluation Metric for Machine Learning Models of Security Applications**," IEEE Networking Letters, Vol. 2, Issue 4, December 2020, pp. 207-211, http://www.cse.wustl.edu/~jain/papers/safety.htm

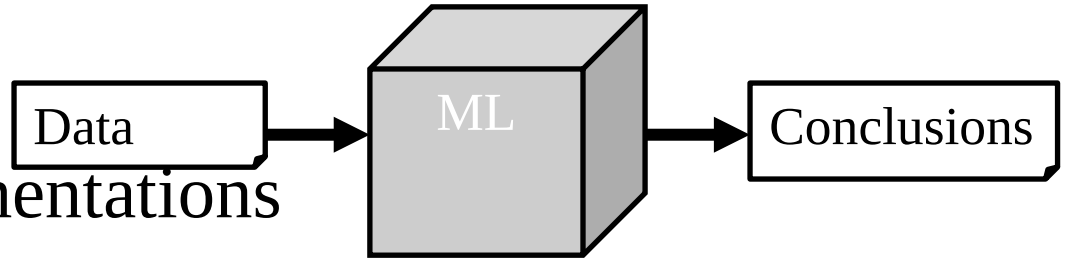
6. Results Not Explainable

- Would you trust AI to diagnose your disease?
- No, because you have no idea of why the results are what they are

Machine Learning is what only machines can do, but human cannot do and cannot explain

- AI is a black box
- Can't discover bugs in ML model implementations
- Need Trustable AI = Explainable AI

  ⇒ Models to explain the AI predictions so that humans can understand



Ref: Maede Zolanvari, Zebo Yang, Khaled Khan, Raj Jain, and Nader Meskin, "TRUST XAI: A Novel Model for Explainable AI with An Example Using IIoT Security," IEEE IoT Journal, preliminary acceptance, September 2021.

Jain's List of ~~Issues~~ Challenges with AI

1. No Domain Expertise
2. Random Datasets
3. Imbalance of Security Data
4. Wrong Performance Metrics
5. Too Few or Too Many Features
6. Results Not Explainable
7. No Sensitivity Analysis
8. No Real-World Validation
9. Omitting Assumptions and Limitations

Issues ⇒ Challenges ⇒ Opportunities for Research.

Blockchains

1. Satoshi Nakamoto invented Bitcoin

2. He used blockchains to make it decentralized
3. Since then blockchains have found numerous other applications
4. Blockchains are distributed, decentralized (no single point of control), trustless (no need to trust other parties), secure (Elliptic Curve Cryptography) with non-repudiation guarantee (all transactions are signed)

Example of a Contract: Wedding

Wedding (Cont)

- **Centralized**

- Centralized registry
- Single point of failure
- Easier to hacked

- Decentralized

- Decentralized
- No single point of failure
- Very difficult to hack

Trend: Centralized → Decentralized

- **Banks**: Allow money transfer between two accounts
- **City Records**: Wedding registers, Property ownership
- **Networks**: Certificate Authorities, DNS

- In all cases:
  - There is a central third party to be trusted
  - Central party maintains a large database ⇒ Attracts Hackers
  - Central party may be hacked ⇒ Affects millions
  - Central party is a single point of failure. Can malfunction or be bribed

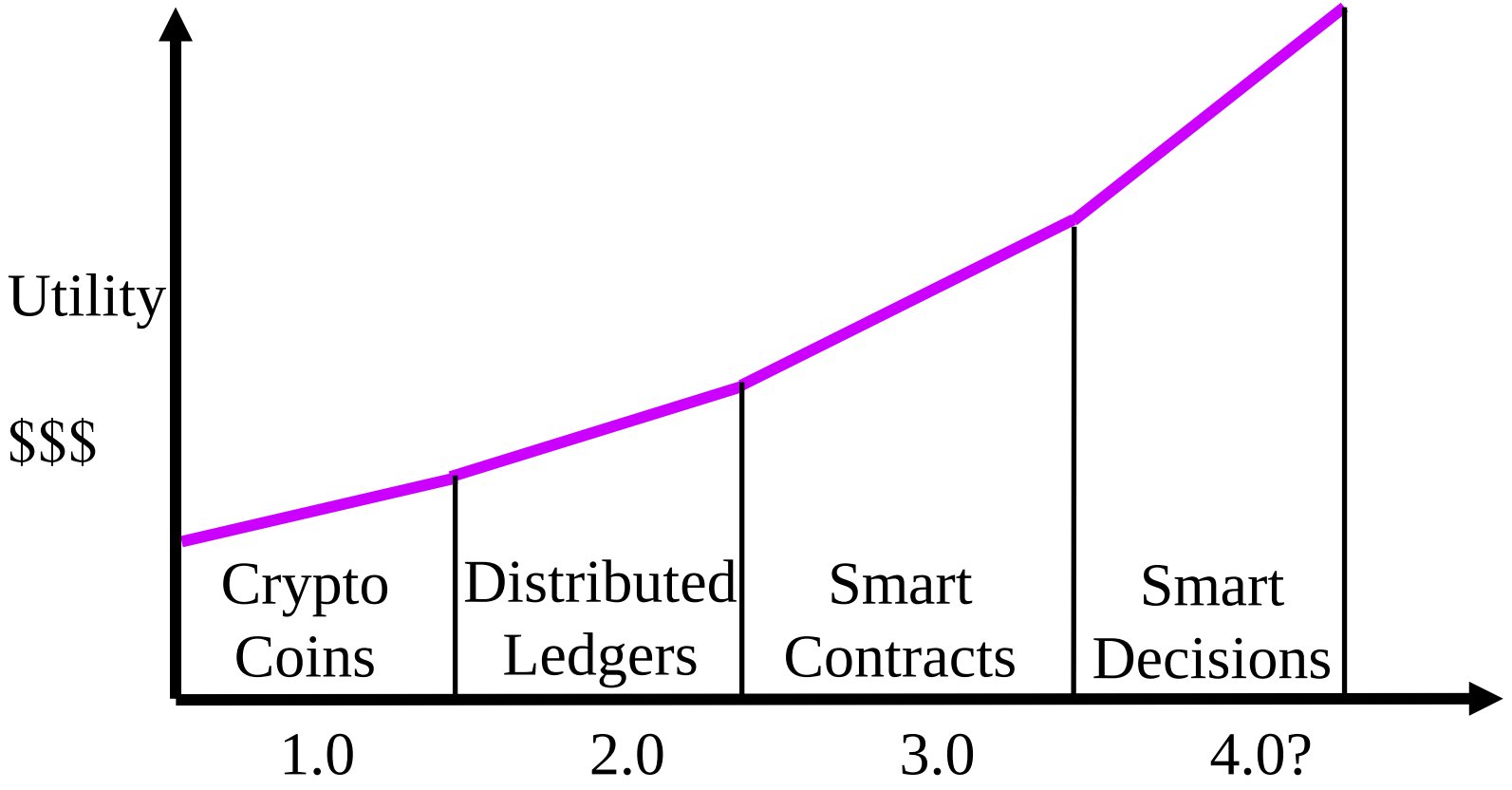
Our Goal

- Moving the chain from deterministic to **probabilistic**
- Moving the chain from storage to **computation**
- Moving the chain from data to **knowledge**
- Moving the chain from information to **decision making**

⇒ A blockchain that provides knowledge ⇒ A **knowledge chain** would be more useful

Ref: Tara Salman, Raj Jain, Lav Gupta, "**A Reputation Management Framework for Knowledge-Based and Probabilistic Blockchains**," IEEE 1st International Workshop on Advances in Artificial Intelligence for Blockchain (AIChain 2019), held in conjunction with the 2019 IEEE International Conference on Blockchain, Atlanta, July 14, 2019, http://www.cse.wustl.edu/~jain/papers/rpmcewa.htm

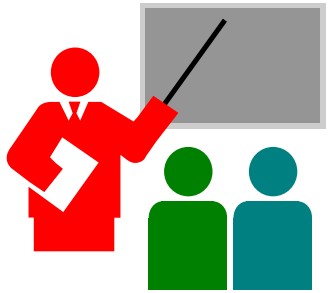
Blockchain Generations



Key Distinction of Our Research

- Goal: Impact to the real-world
  DECbit congestion indication in almost all networking architectures since its invention
- Funded by industry partners:
  Intel, Cisco, Broadcom, Boeing, …
- Impact real-world by participating in standards organizations and industry forums:
  ATM Forum, IEEE Standards, American National Standards Institute (ANSI), Internet Engineering Task Force (IETF), WiMAX Forum
- Work on long term as well as short term research

Summary



1. AI → Explainable AI → Federated AI
2. Blockchains
   → Blockchains with AI = Knowledge Chains
3. IoT Intelligent IoT Secure and Intelligent IoT
   - Industrial systems
   - Medical systems
   - Agriculture Systems
4. Research for Impact

References: Class Recordings

- Recordings of all of my classes and talks are available on YouTube and on my website:

1. CSE 473: Introduction to Computer Networks, http://www.cse.wustl.edu/~jain/cse473-21/index.html
2. CSE 571S: Network Security, http://www.cse.wustl.edu/~jain/cse571-17/index.html
3. CSE 574S: Wireless Networks, http://www.cse.wustl.edu/~jain/cse574-20/index.html
4. CSE 567: Computer Systems Analysis http://www.cse.wustl.edu/~jain/cse567-17/index.html
5. CSE 570: Recent Advances in Networking http://www.cse.wustl.edu/~jain/cse570-19/index.html

Our Publications on AI

- L. Gupta, T. Salman, A. Ghubaish, D. Unal, A. K. Al-Ali, R. Jain, "Cybersecurity of multi-cloud healthcare systems: A hierarchical deep learning approach," Applied Soft Computing (2022), 5 January 2022, http://www.cse.wustl.edu/~jain/papers/muse.htm
- Maede Zolanvari, Ali Ghubaish, and Raj Jain, "ADDAI: Anomaly Detection using Distributed AI," in Proceedings of IEEE ICNSC (International Conference on Networking, Sensing and Control), October 2021., http://www.cse.wustl.edu/~jain/papers/addai.htm
- M Zolanvari, Z Yang, K Khan, R Jain, N Meskin, "TRUST XAI: Model-Agnostic Explanations for AI With a Case Study on IIoT Security," IEEE Internet of Things Journal, 2021, http://www.cse.wustl.edu/~jain/papers/trustxai.htm
- Tara Salman, Ali Ghubaish, Devrim Unal, Raj Jain, "Safety Score as an Evaluation Metric for Machine Learning Models of Security Applications," IEEE Networking Letters, Vol. 2, Issue 4, December 2020, pp. 207-211, http://www.cse.wustl.edu/~jain/papers/safety.htm
- Lav Gupta, Tara Salman, Maede Zolanvari, Aiman Erbad, Raj Jain, "Fault And Performance Management In Multi-Cloud Virtual Network Services Using AI: A Tutorial And A Case Study," Computer Networks, Pre-Proof published on 14 Oct 2019, http://www.cse.wustl.edu/~jain/papers/fp_comst.htm

Our Publications on AI (Cont)

- Lav Gupta, Tara Salman, Ria Das, Aiman Erbad, Raj Jain, Mohammed Samaka, "HYPER-VINES: A HYbrid Learning Fault and Performance Issues ERadicator for VIrtual NEtwork Services over Multi-cloud," International Workshop on Computing, Networking and Communications (CNC19) at the International Conference on Computing, Networking and Communications (ICNC 2019), Honolulu, Hawaii, Feb. 18-21, 2019, 7 pp., http://www.cse.wustl.edu/~jain/papers/hypervin.htm
- Maede Zolanvari, Marcio A. Teixeira, Raj Jain, "Effect of Imbalanced Datasets on Security of Industrial IoT Using Machine Learning," 2018 IEEE International Conference on Intelligence and Security Informatics (ISI), Miami FL, Nov. 9 - 11, 2018, 6 pp., http://www.cse.wustl.edu/~jain/papers/imb_isi.htm
- Tara Salman, Deval Bhamare, Aiman Erbad, Raj Jain, Mohammed Samaka, "Machine Learning for Anomaly Detection and Categorization in Multi-cloud Environments," The 4th IEEE International Conference on Cyber Security and Cloud Computing (IEEE CSCloud 2017), New York, June 26-28, 2017, http://www.cse.wustl.edu/~jain/papers/cscloud.htm

Our Publications on AI (Cont)

- Deval Bhamare, Tara Salman, Mohammed Samaka, Aiman Erbad, Raj Jain, "Feasibility of Supervised Machine Learning for Cloud Security," 3rd International Conference on Information Science and Security (ICISS2016), December 19th - 22nd, 2016, Pattaya, Thailand, http://www.cse.wustl.edu/~jain/papers/iciss16.htm

Our Publications on ICS Security

- M Elnour, N Meskin, K Khan, R Jain, "Application of data-driven attack detection framework for secure operation in smart buildings," Sustainable Cities and Society 69, 102816, 2021, http://www.cse.wustl.edu/~jain/papers/secbldg.htm
- M Elnour, N Meskin, K Khan, R Jain, "HVAC System Attack Detection Dataset," Data in Brief, 107166, 2021, http://www.cse.wustl.edu/~jain/papers/hvac.htm
- MA Teixeira, M Zolanvari, KM Khan, R Jain, N Meskin, "Flow-based intrusion detection algorithm for supervisory control and data acquisition systems: A real-time approach," IET Cyber-Physical Systems: Theory & Applications, 2021, http://www.cse.wustl.edu/~jain/papers/ids_ijis.htm
- M. Elnour, N. Meskin, K. Khan, R. Jain, "A Dual-Isolation-Forests-Based Attack Detection Framework for Industrial Control Systems," IEEE Access, Vol. 8, 19 February 2020, pp. 36639 - 36651, http://www.cse.wustl.edu/~jain/papers/dif.htm
- Mariam Elnour, Nader Meskin, Khaled Khan, Raj Jain, "A Dual-Isolation-Forests-Based Attack Detection Framework for Industrial Control Systems," IEEE Access, Vol. 8, 19 February 2020, pp. 36639 - 36651, http://www.cse.wustl.edu/~jain/papers/dif.htm

Our Publications on ICS Security

- *Mariam Elnour, Nader Meskin, Khaled M. Khan, Raj Jain, Syed Zaidi, Hammadur Siddiqui, "Full-Scale Seawater Reverse Osmosis Desalination Plant Simulator," 21st IFAC World Congress in Berlin, Germany, July 12-17, 2020, 8 pp., http://www.cse.wustl.edu/~jain/papers/swrodp.htm*
- *D. Bhamare, M. Zolanvari, A. Erbad, R. Jain, K. Khan, N. Meskin, "Cybersecurity for Industrial Control Systems: A Survey," Computers and Security, Elsevier, Volume 89, February 2020, Article 101677, http://www.cse.wustl.edu/~jain/papers/ics_survey.htm*
- M. Zolanvari, M. Teixeira, L. Gupta, K. Khan, R. Jain, "Machine Learning Based Network Vulnerability Analysis of Industrial Internet of Things," IEEE Internet of Things Journal, Vol. 6, Issue 4, Aug 2019, http://www.cse.wustl.edu/~jain/papers/vulnerab.htm
- M. Zolanvari, M. Teixeira, R. Jain, "Effect of Imbalanced Datasets on Security of Industrial IoT Using Machine Learning," 2018 IEEE International Conference on Intelligence and Security Informatics (ISI), Miami FL, Nov. 9 - 11, 2018, 6 pp., http://www.cse.wustl.edu/~jain/papers/imb_isi.htm
- Marcio Andrey Teixeira, Tara Salman, Maede Zolanvari, Raj Jain, Nader Meskin, and Mohammed Samaka, "SCADA System Testbed for Cybersecurity Research Using Machine Learning Approach," Future Internet 2018, 10(8), 76, http://www.cse.wustl.edu/~jain/papers/ics_ml.htm

Our Publications on Healthcare Security

- Ali Ghubaish, Tara Salman, Maede Zolanvari, Devrim Unal, Abdulla Khalid Al-Ali, Raj Jain, "Recent Advances in the Internet of Medical Things (IoMT) Systems Security," IEEE Internet of Things Journal, Vol. 8, Issue 11, June 1, 2021, http://www.cse.wustl.edu/~jain/papers/iomt_iot.htm
- Anar A. Hady, Ali Ghubaish, T. Salman, Devrim Unal, and R. Jain, "Intrusion Detection System for Healthcare Systems Using Medical and Network Data: A Comparison Study," IEEE Access, June 2020, http://www.cse.wustl.edu/~jain/papers/hms.htm

Our Publications on Blockchains

- Tara Salman, Maede Zolanvari, Aiman Erbad, Raj Jain, and Mohammed Samaka, "Security Services Using Blockchains:A State of the Art Survey" IEEE Communications Surveys and Tutorials, First Quarter 2019, Volume 21, Issue 1, 858-880 pp., http://www.cse.wustl.edu/~jain/papers/bcs.htm
- Tara Salman, Raj Jain, Lav Gupta, "A Reputation Management Framework for Knowledge-Based and Probabilistic Blockchains," IEEE 1st International Workshop on Advances in Artificial Intelligence for Blockchain (AIChain 2019), held in conjunction with the 2019 IEEE International Conference on Blockchain, Atlanta, July 14, 2019, http://www.cse.wustl.edu/~jain/papers/rpmcewa.htm
- Tara Salman, Raj Jain, and Lav Gupta, "Probabilistic Blockchains: A Blockchain Paradigm for Collaborative Decision-Making," 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON 2018), New York, NY, November 8-10, 2018, 9 pp., http://www.cse.wustl.edu/~jain/papers/pbc_uem.htm

Acronyms

- 3GPP Third Generation Partnership Project
- AI Artificial Intelligence
- ANSI American National Standards Institute
- AT&T American Telephone and Telegraph
- BSS Business Support Services
- CA California
- CGNAT Carrier Grade Network Address Translator
- CSE Computer Science and Engineering
- DECbit Digital Equipment Corporation Bit
- IEEE Institution of Electrical and Electronic Engineering
- IoT Internet of Things
- ML Machine Learning
- MO Missouri
- MS Master of Science
- NFV Network Function Virtualization
- NTT Nippon Telephone and Telegraph

Acronyms (Cont)

- OpenADN Open Application Delivery Networking
- OSS Operations Support Services
- SON Self-Organizing Networks
- TV Television
- UK United Kingdom
- US United States
- VC Venture Capital
- WAN Wide Area Network
- WiMAX Worldwide Interoperability for Microwave Access
- WUSTL Washington University in St. Louis

Scan This to Download These Slides

Raj Jain
Rajjain.com

http://www.cse.wustl.edu/~jain/talks/cs59122.htm

Our Courses on YouTube

CSE567M: Computer Systems Analysis (Spring 2013),
https://www.youtube.com/playlist?list=PLjGG94etKypJEKjNAa1n_1X0bWWNyZcof

CSE473S: Introduction to Computer Networks (Fall 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypJWOSPMh8Azcgy5e_10TiDw

CSE 570: Recent Advances in Networking (Spring 2013)
https://www.youtube.com/playlist?list=PLjGG94etKypLHyBN8mOgwJLHD2FFIMGq5

CSE571S: Network Security (Fall 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypKvzfVtutHcPFJXumyyg93u

Video Podcasts of Prof. Raj Jain's Lectures,
https://www.youtube.com/channel/UCN4-5wzNP9-ruOzQMs-8NUw