ICN-Enabled Secure Edge Networking with Augmented Reality (ICE-AR): A Quick Overview

ICN-WEN ANNUAL MEETING

JUNE 19 2018

From the ICN-WEN solicitation

Vison

- "The wireless edge-based nature of the B5G applications provides an opportunity to overcome these challenges by deploying 'clean-slate' approaches as self-contained information 'islands'.
- "Within the island, an information-centric infrastructure labels, stores, retrieves and routes information among the devices, data repositories, and local compute and analytics systems.

Validation

- "It is expected that research proposals will include *a strong prototyping component*.
- "In addition to proof-of-concept, prototypes help to quantify approach benefits and effectiveness, and to better understand implementation and deployment challenges within *more realistic contexts*."

ICE-AR: our starting point

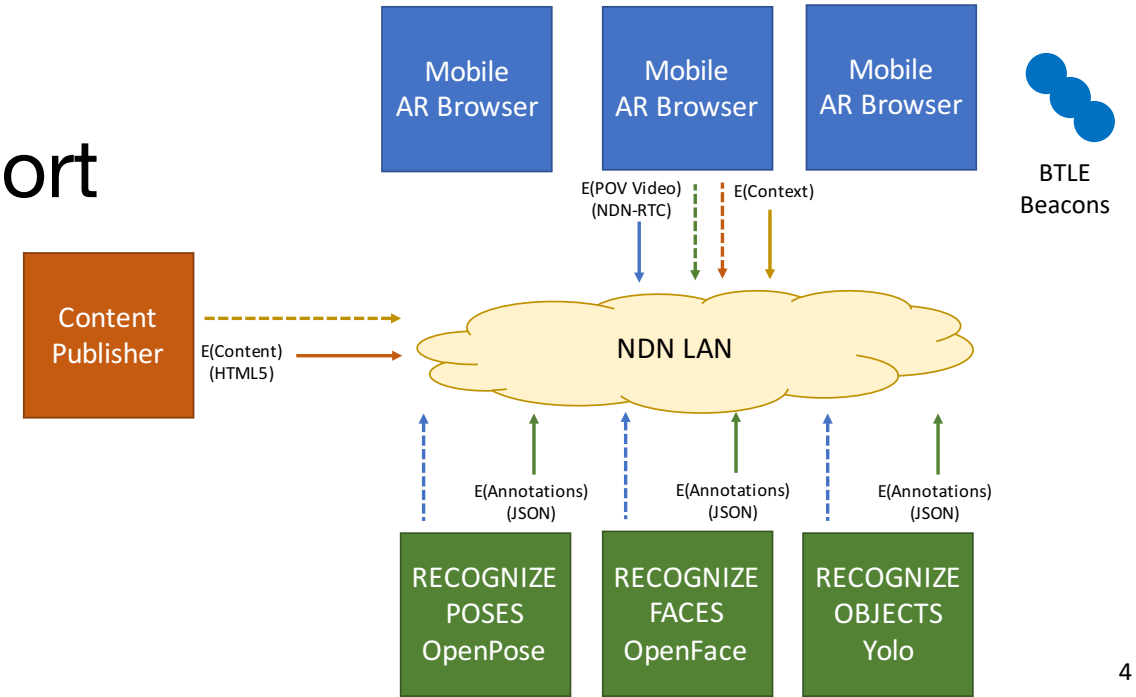
The NDN design and implementation from NSF FIA program:

- Laid the architectural foundation
- Provided the initial system codebase
  - enabling us to tackle the challenge of supporting AR apps over wireless edge from day one
- Established proven research methodology: application-driven architecture development

ICE-AR started with AR app prototyping

Driving research along multiple dimensions

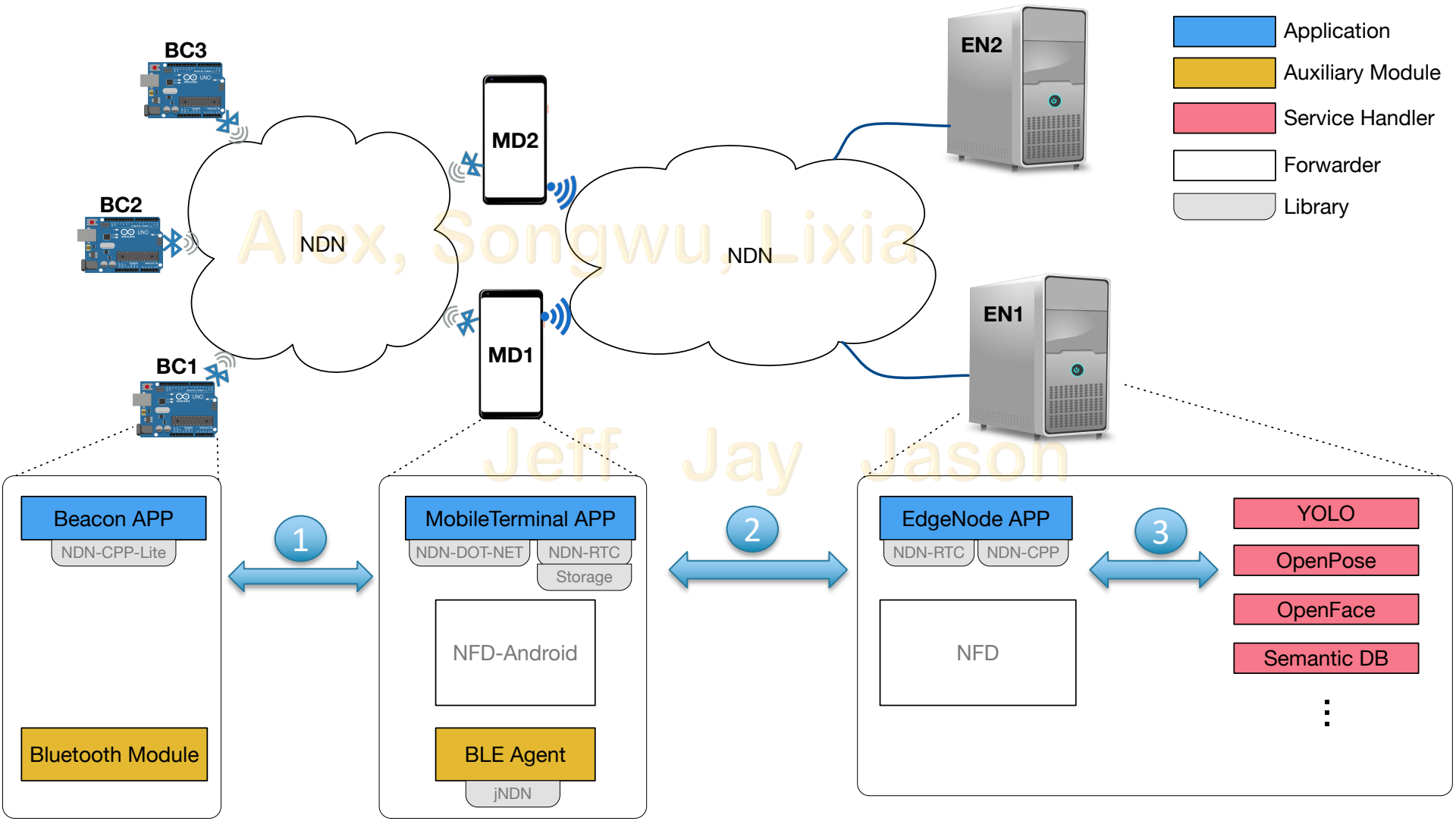
- ◊ Data naming
- ◊ Acceleration as a service
- ◊ Local resource discovery
- ◊ Security
- ◊ Lower layer support



The people and process

- Jeff: leading the app development
- Jason / Jessie: acceleration
- Jay / Alex: security
- Songwu/Yuanjie: wireless
- Alex / Yanbiao: NDN architecture, protocol, security, and codebase

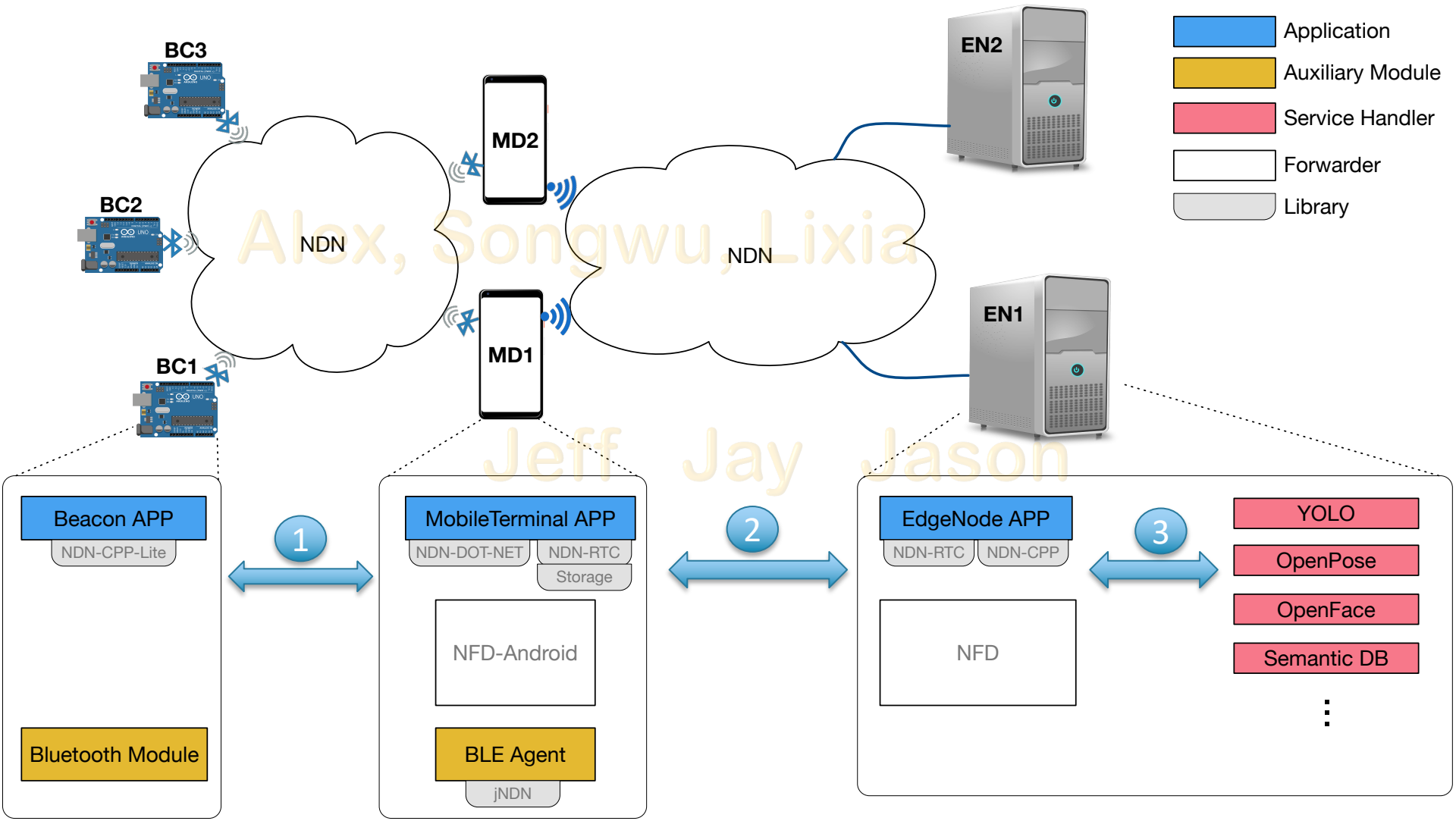
Interconnecting pieces into a whole



Three aspects of the integration

1. Bluetooth low energy beaconing
   - BCs advertise locations
   - MDs monitor BC presence/absence, retrieve location-based context
2. Edge service discovery (in progress)
   - ENs announce service prefixes, and/or
   - MDs discovery services with self-learning
3. Acceleration as a service
   - edge node apps register service name with local NFD
   - NFD dispatches incoming requests (for service) by their names
     - now: use file pipes
     - Next step: use native NDN Interest/Data exchange

Interconnecting pieces into a whole



Application aspects of integration

- NDN-RTC library integration for mobile platforms (Android)
  - made library compatible for Android cross-compilation
- Developer-friendly approach for fetching data using C# CNL
  - generalized objects fetching, abstracting away from interest-data exchange
- Local storage for published NDN data packets (video stream)
- Edge node containerization
  - Containerize each processing module as a Docker container for easier and faster deployment
- AR functionality
  - Semantic DB
    - storing annotations for later retrieval of “the most similar scenes”
  - spatially-aware rendering of bounding boxes over live video, and tracking bounding boxes from frame to frame (Kalman filter)

Accomplishment – AR app

- Design and prototype of end-to-end AR application over NDN.
  - mobile device (MD) publish POV video, stores in a local NDN repo
  - Edge service nodes (EN) consume video, generate/publish annotations using multiple ML modules.
  - MD fetches the resulting context annotations, re-associated with the frame (all in perceptual real-time for the end-user)
  - integration of BTLE beacons provides additional context for frames
  - Continuous query of an edge database to find matching content
- Provided acceleration workload, security, and forwarding requirements to the rest of the team
- Feedback/refinement to NDN development
  - Further developed the NDN Common Name Library
    - enable app developers to interact with NDN data without handling Interest/Data exchange details.
  - Improved real-time video conf. library (NDN-RTC) to support AR app
  - following the latest NDN protocol refinement

Accomplishment – Acceleration

- Objectives
  - Use FPGA in the edge node for OpenPose acceleration.
  - Real-time processing with low energy cost.
- Achievements
  - Conducted the study of network compression
    - Compressed the network size by 70% with 2% of accuracy loss
  - Designed and implemented the FPGA accelerator for OpenPose algorithm based on the reduced network.
  - The current solution is estimated to achieve 3.2 FPS with 4.7x better energy efficiency than GPU (Nvidia GPU).

Accomplishment – security development

- Creating a trust framework for diverse types of users in a campus setting
  - faculty members
  - students
  - visitors
- Exploring mechanisms for access control for AR applications
  - sharing content between mobile devices and edge servers
- Leveraging existing NDN trust and access control solutions, refine/enhance to meet ICN-WEN needs

Accomplishment – wireless codesign

- Overcome the barriers of current wireless codesign
  - Performance optimizations as the main goal for codesign
  - Limited solution space: typical two layers for cross-layer design (e.g., transport & link, network & link)
  - Blackbox/graybox based design: Cannot reason whether it works or not in practice
- Preliminary experimental characterizations
  - AR/VR network traffic patterns, and
  - Deficiencies of 4G wireless for AR/VR [SIGMETRICS'18]
- NDN-based approach to wireless codesign
  - Top-down rather than bottom-up: from communication-driven wireless networking to app-driven networking
  - Sharing app namespace across layers: enabler for cross-layer optimization
    - Cover both system performance and reliability aspects
  - Built-in analytics via NDN to learn what and reason why
  - Identify the opportunities of wireless-aware NDN forwarding strategy, and NDN-aware wireless optimization

What we are learning from experimentation

What we are learning from experimentation: NDN and AR

- AR-over-ICN motivates integration of multiple data models at app/service layer(s)
  - Today: one has to pick from DOM, media object/stream, or scene graph data models for content.
    - Makes it difficult to choose one prototyping platform: Unity has best AR integration for prototyping, but limited platform for building a "browser." Chromium may be best option, but heavyweight development and limited data model integration.
- Need more general work on advanced NDN namespace design concepts, to open up namespace design to other domains/specializations
  - Generally, the next step / bottleneck in integrating other work is developing complex namespaces for the components, e.g. acceleration-as-a-service and security integration.
  - First pass can build from existing approaches but little written about integrating namespace design considerations.

What we are learning from experimentation: NDN and edge computing

- Satya@CMU: "The view of the cloud as a single, logically centralized and managed entity that serves as a computing utility simplified questions relating to *resource discovery*, *trust*, and *business models*."
- Handle the three issues in edge computing: using NDN as the architectural base
  - Routing/forwarding by app names →local resource discovery at network layer
  - Trust and security: via NDN security framework
  - Business model: yet to be developed; the best way forward: enabling open collaboration & competition

Obstacles encountered

- The existing legacy systems and lower layer communication supports present a major barrier to experimentation with new architecture
  - Existing system platforms do not allow direct access to lower layers
  - Existing wireless protocols do not support broadcast/multicast, or support efficiently
- in playing with Bluetooth low energy
  - System restrictions, e.g., libraries vary in different platforms; Android 8's BLE library has issues in some phones
  - BLE mesh networking seems more friendly to NDN, but lacks of documentations

Evaluation: what/how to measure

From the ICN-WEN solicitation:

- “Investigate and prototype architectures, protocols, and/or technologies based on the application of ICN to an "island" wireless edge network, and
- “demonstrate (quantifiably) the effectiveness in addressing the challenges of B5G applications as compared to non-ICN approaches.”

Evaluation at two levels:

- Qualitative benefits *In progress*
  - programmability, flexibility of application, ease of management, usability when applied at scale, robustness against attacks or failures, extensibility, composability, and so on.
- Quantitative performance metrics should be used to establish the benefits of a proposed scheme in the context of wireless network and device research *Facing challenges*
  - Lack of access to lower layers
  - Existing wireless protocols unfit to ICN
  - No plan on device research

Application – goals update

- Spring/summer: Finish current ICE-AR prototype and browser design.
  - move towards driver application context-content exchange.
  - Update protocol for real-time media fetching.
  - Complete AR experience for the public. (with other support)
  - Incorporate NDN platform advances in mobility, autoconfiguration
  - Integrate initial security, acceleration, wireless work
- Fall: Start on browser generalization.
  - Just-in-time code delivery.
  - Acceleration-as-a-service.
  - Leverage lower-layer capabilities for low latency and heterogeneous wireless support.
  - Generalization of the web: context-content exchange (with other support).

Wireless Codesign – goals update

- Summer: Complete first-version codesign
  - Complete analytics design
  - First prototype over WiFi
- Fall: Prototype of the first-version design with new design for
  - mobility case
  - codesign with security in the loop
- Winter: Second-version prototype
  - Validation of the design through experiments
  - Refine design with IoT in mind

Networking & Acceleration – goals update

Summer + fall 2018:

- "must shift impl. to assertive, opportunistic connectivity with minimal developer/deployer/ user intervention"
  - "at this time NDN platform and forwarding require too much careful setup for each circumstance"
  - "need integration with mobility support and NDN-over-wireless"
- Acceleration: test and integrate the FPGA kernel with NDN framework.
  - Explore other compression approaches (e.g., sparse pruning, quantization) to reduce the computation cost.

Summarizing: what we have learned

- named, secured data enables integration of networking, storage, and processing into a coherent system through
- App-driven architecture development works

Why apps as the driver

- identifies all necessary supporting functions
- sees the interactions/dependencies
- offers insight into system-wide optimization
- exposes all obstacles in new architecture realization

In contrast to bottom-up approach

- Seeing pieces in isolation
  - looking at one piece at a time
  - difficult to see the overall picture
- Local optima (algorithmic tuning)

What we have learned

- named, secured data enables integration of networking, storage, and processing into a coherent system through
- App-driven architecture development works
- Main challenge: NDN namespace ties together app, network, and security
  - Bringing the power of ICN
  - Making the design of each piece more challenging
- Ongoing effort: extracting general design guidelines through more experimentations

Takeaway

Question from Intel: if ICN were to provide functions and services in the network based on naming, how would it work?

from ICE-AR proposal:

- Innovative naming & name discovery design, enabling discovery and forwarding of content data from diverse users, offering low-latency content retrieval, and seamlessly embeds edge computing.
- NDN-enabled edge computing, which supports heterogeneous computing hardware and offers the novel abstraction of “Accelerator-As-A-Service."
- Scalable schematized trust management and data-centric security, which provides security as a built-in feature rather than an after-thought patch.

Recent publications

1. “DICE: Dynamic Multi-RAT Selection in the ICN-enabled Wireless Edge”, MobiArch Workshop, July 2017.
2. “LASeR: Lightweight authentication and secured routing for NDN IoT in smart cities”, IoT Journal, Feb 2018.
3. “Achieving Resilient Data Availability in Wireless Sensor Networks”, ICN-SRA Workshop, May 2018
4. “Supporting Mobile VR in LTE Networks: How Close Are We?” ACM SIGMETRICS, June 2018.
5. “Towards Edge Computing Over Named Data Networking”, IEEE International Conference on Edge Computing, July 2018.
6. “TACTIC: Tag-based Access ConTrol Framework for the Information-Centric Wireless Edge Networks” IEEE ICDCS, July 2018.
7. “Mobile Data Repositories at the Edge”, USENIX HotEdge Workshop, July 2018.
8. “Supporting Augmented Reality: Looking Beyond Performance”, ACM SIGCOMM 2018 Workshop on VR/AR Network, August 2018.
9. “Real-Time Data Discovery In Named Data Networking”, IEEE HotICN, August 2018
10. “AccConF: An access control framework for leveraging in-network cached data in the ICN-enabled wireless edge” Transactions on Dependable and Secure Computing, in press
11. “An Overview of Security Support in Named Data Networking”, IEEE Communication Magazine, special issue on ICN Security

Papers under submission

- "Distributed Dataset Synchronization in Mobile Ad Hoc Networks over NDN"
- "KITE: Producer Mobility Support in Named Data Networking"
- "NDN Host Model"

Tech reports

- "NDN Automatic Prefix Propagation" NDN-0045, Feb 2018.
- "VectorSync: Distributed Dataset Synchronization over Named Data Networking" NDN-0056, March 2018.
- "NDN Device Secure Sign-On Protocol", in progress.
- "NDN Host Multihoming", in progress.

Verizon Summary

- service/info centric network happening now
- multitudinous edge access and clouds under mobility are challenging
- deployabiligy and service quality assurance are critical for ROI and user experience

Instructions from Intel

- Overview presentations - executive summaries of
  - what each team has done
  - putting the work of each PI in some overall context and how it relates to ICN
- Session presentations – 15min targeted presentations. The session organizer will lead a 45-minute discussion afterwards. The session organizers want a couple bullet points/sentences on what you will be talking about in order to prepare for the discussion.
- Intel questions for the architecture session
  - If ICN were to provide functions and services in the network based on naming, how would it work? And if it does, where does the computation take place? Does the consumer ask for function and provide the data? How does this affect the security of the data and who owns the transformed data and how?
  - Given ICN is more reliant on next hop or nearest hop where there might be content cached, does it need to be architected to take into account different wireless schemes (contention-based, contention-free, licensed vs. unlicensed?).
  - What is the wireless edge for ICN-WEN? Is it the AP or 4G/5G Base station or is it just simply the first hop from the producer? Is it where the wireless joins the wired?

Outline

- Solicitation: prototyping a new architecture
- FIA results as starting point
- Why app-driven: optimization by architecture
- Achievements
- What we have learned
- Bring up big questions at end of my talk
  - where we are heading to
  - Relation with edge computing
  - Business model

From ICE-AR Proposal

- We believe that NDN can enable AR at wireless edge through systematic naming and retrieving named data at network layer.
- NDN design principles emphasize
  - application-level data naming and data-centric security
  - asynchronous publishing and consumption, and
  - enabling efficient use of local and proximate resources.
- We will apply them to build a system that
  - unifies the latest advances in wireless communication with domain specific computing technologies to accelerate AR at wireless edge, and
  - deliver robust performance for large groups of people interacting in real-time with data and content.

Jeff's talk on app

- Recap of ICE-AR application vision.
- Current prototype status.
  - Video stream published from mobile and stored in a local repo on the device
  - Video consumed by edge services provide "deep context" via ML analysis of stream and stores in a local repo
  - Mobile consumes edge-published context for current frame and renders
  - Mobile also issues a periodic query to the edge DB (currently over http) for frames similar to the current one
  - Best match historical frames are requested using random-access retrieval via NDN names
- Next steps
  - Integrate identity management being prepared by network group
  - Finish edge-as-a-service design
  - Expand work on how to name context
  - Integrate data-centric security from NMSU, edge acceleration from Jason Cong's group

Making Broader Impact

(quoted from ICE-AR proposal)

- If successful, this project will be among the first that *integrates ICN, wireless edge, and edge computation*.
- It will address critical challenges in the deployment of *realistic AR applications at scale*.
- This project will involve both under- and graduate students, allowing them to gain understanding of networking at the architectural level.

Testbed and evaluation

- Complete ICE-AR testbed
  - Security support implementation
  - Mobility support implementation
- UCLA testbed
- NMSU?
  - featuring end user devices, vehicles, and drones, all equipped with NDN protocols and instrumented with latest wireless equipment, as well as sensors
- Interconnection of the two? When?
- Emulations and simulations ?

- Ongoing proof-of-concept prototype
  - How our approach further reduces AR latency over multiple WiFi interfaces
    - Application-guided retransmissions
    - Chase for latency rather than highest speed for AR context content transfer
    - Reason wireless-related failures
    - Towards predictive design during mobility