SafEdge for Residential Networks

Privacy from the Bottom Up

Ph.D. Thesis Proposal

Aldo Cassola
College of Computer and Information Science
Northeastern University

Committee:
David Choffnes
Alan Mislove
Guevara Noubir
Omprakash Gnawali (U. of Houston)

February 12 2014

Trends in Mobile Networks

- Internet is increasingly mobile. According to [CISCO2013]:
- Mobile data volume grew 70%, (500 Petabytes/month)
- Smartphones: 92% of handset traffic

- Connected mobile tablets online increased 2.5x (36M)
- Network speed, more than doubled
- Over 30% of traffic is offloaded to Femtocell or Wi-Fi, expected to increase [DeviceScape] [3GPP TS 23.261]

The Era of Free Cloud Services

- Increased connectivity: users expect ubiquitous access
  - Providers struggle to deliver large volumes, reduce cell sizes, offload to Wi-Fi
- Offerings for file sharing and synchronization
  - Dropbox (200M users), Google docs (120M), Microsoft SkyDrive (250M)
- Email, communications, streaming
  - Gmail (425M users), Hotmail (420M), Skype (660M), Youtube (1B)
- Social Networks
  - Twitter (218M), Facebook (1B)

- **What are the privacy implications?**

Security and Privacy Concerns

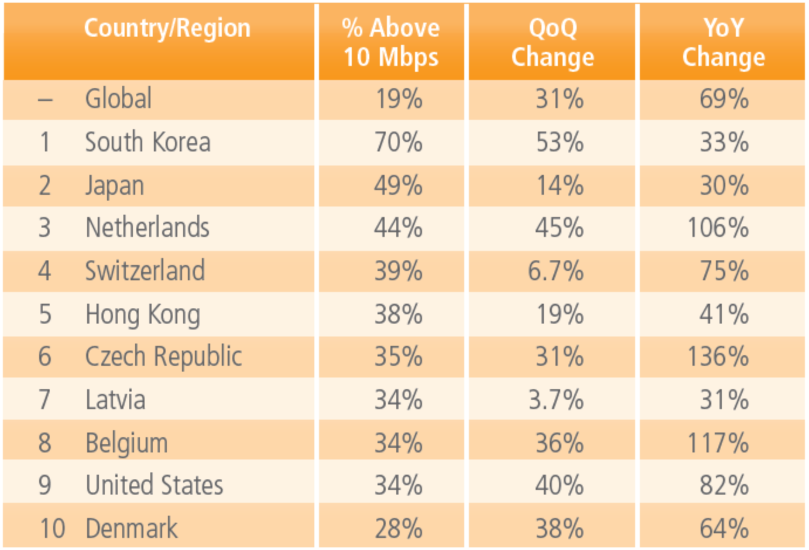
- Network access:
  - Mobile Network operators can access handset data and location
  - Offloading to Open Wi-Fi APs encourages AP impersonation (Evil Twins, credential hijacking)

- Data protection:
  - Free services like plaintext data (plaintext Gmail → Ads)
  - Clients may snoop into data (Skype visiting “encrypted” URLs)
  - Encrypted data access can leak information

- User Tracking:
  - Application providers can infer personal information from usage (e.g. weekday usage leaks workplace)

The Residential Space

- Network providers try to bring the network closer to users
- Deployment is hard and expensive
- Residential Broadband continues growth [AkamaiSOTI 2014, PEWINT2013]
- Residential devices: always on, capable, low failure rate (10K hours)

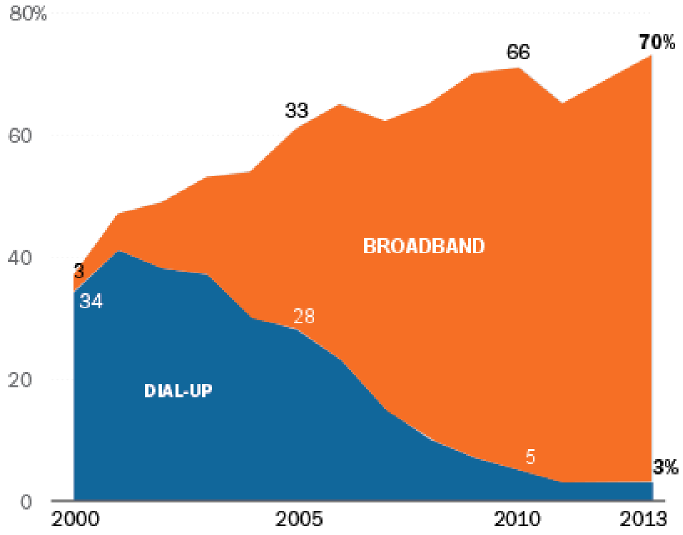


	Country/Region	% Above 10 Mbps	QoQ Change	YoY Change
–	Global	19%	31%	69%
1	South Korea	70%	53%	33%
2	Japan	49%	14%	30%
3	Netherlands	44%	45%	106%
4	Switzerland	39%	6.7%	75%
5	Hong Kong	38%	19%	41%
6	Czech Republic	35%	31%	136%
7	Latvia	34%	3.7%	31%
8	Belgium	34%	36%	117%
9	United States	34%	40%	82%
10	Denmark	28%	38%	64%

Figure 16: High Broadband (>10 Mbps) Connectivity

Home Broadband vs. Dial Up, 2000-2013

Percentage of American adults 18 years and older who access the internet via ...



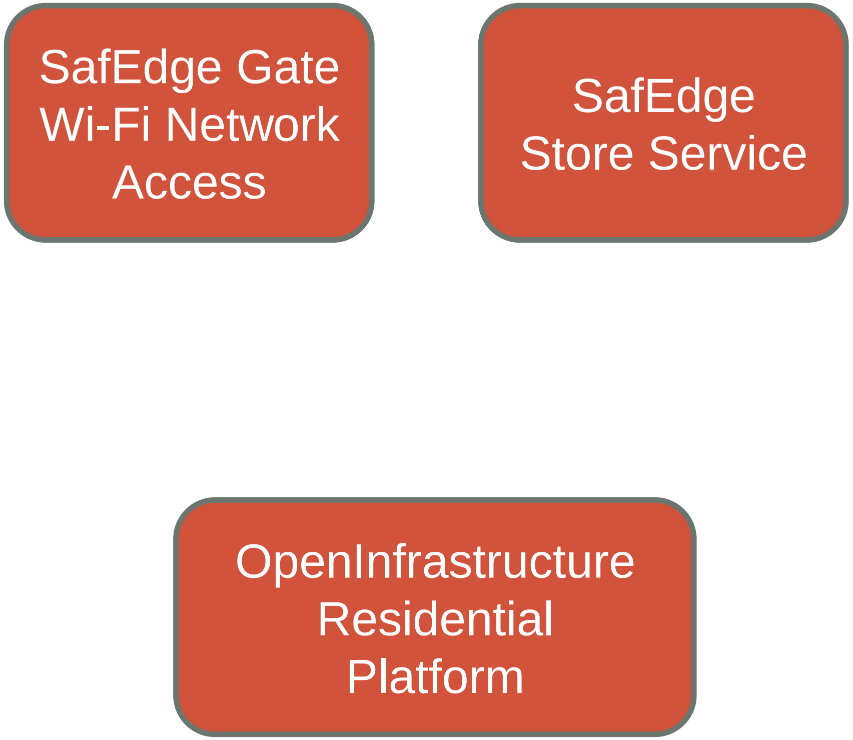
Thesis Statement

Residential Broadband Network access and infrastructure is a suitable bedrock to build network access and cloud services that are at the same time efficient, secure and privacy-protecting.

Focus of this Work

- Contributions:
  - Development and deployment of platform to study residential broadband
  - Identified potential for impersonation in advanced Wi-Fi technologies, and proposed solutions
  - Building new classes of service for more private network access

- 3 Main areas of work:



SafEdge Gate Wi-Fi Network Access

SafEdge Store Service

OpenInfrastructure Residential Platform

Focus of this Work

- **Study Residential Infrastructure**
  - Low-end devices
  - Heterogeneous platforms
  - Limited uplink
  - ➢ **Research and Deployment Platform: OpenInfrastructure**

- **Extend network coverage to smartphones by allowing AP owners to offer backhaul**
  - Home AP owners share network privately
  - Improve network coverage with Wi-Fi
  - ➢ **Access Control and Privacy: SafEdge Gate**

- **Build cloud services running on the Edge: Storage**
  - Integrate privacy protection to service
  - Maximize performance over anonymity networks
  - Minimum impact to existing traffic
  - ➢ **Minimize exposure to service providers: SafEdge Store**

Overview

1. Open Infrastructure
2. Residential Network Access
3. Edge Storage
4. Schedule
5. Questions

Open Infrastructure Testbed

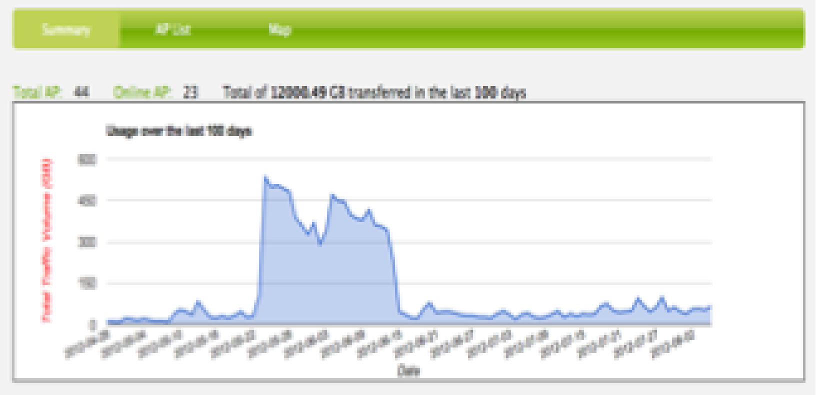
- Suite of hardware and management tools for residential devices
  - Deploy and host new applications and experiments
  - Gather and analyze experiment data
  - Manage devices

- Goal: Offer a homogeneous platform for residential deployments
  - Other testbeds run on well-provisioned networks (PlanetLab)
  - Residential networks are unique (asymmetric, bandwidth- and hardware-limited)
  - First-hand data on usage and connectivity

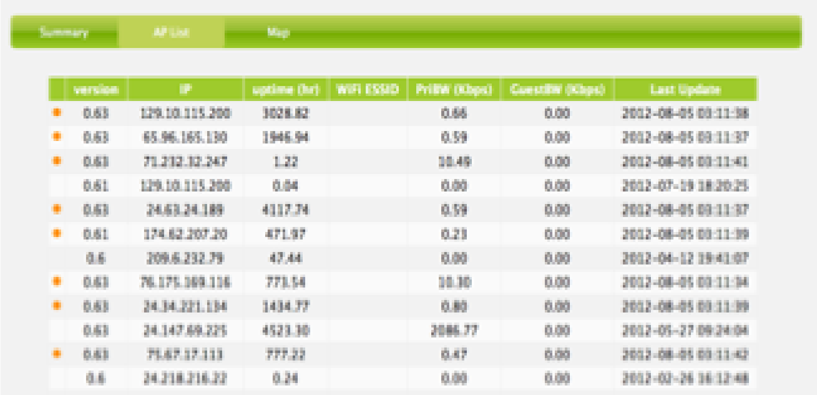
Open Infrastructure Testbed

- Customized OpenWrt software
  - Suite of management and data gathering tools
  - Health and bandwidth capacity monitor
- 802.11n Devices
- 16GB USB flash
- 64MB RAM, 32MB on-board flash, 400MHz CPU
- Web-management Portal

OpenWrt
Wireless Freedom



Total AP: 44 Online AP: 23 Total of 12000.49 GB transferred in the last 100 days



	version	IP	uptime (hr)	WiFi ESSID	PriBW (Kbps)	GuestBW (Kbps)	Last Update
●	0.63	129.10.115.200	3028.82		0.66	0.00	2012-08-05 03:11:38
●	0.63	65.96.165.130	1946.94		0.59	0.00	2012-08-05 03:11:37
●	0.63	71.232.32.247	1.22		10.49	0.00	2012-08-05 03:11:41
	0.61	129.10.115.200	0.04		0.00	0.00	2012-07-19 18:20:25
●	0.63	24.63.24.189	4117.74		0.59	0.00	2012-08-05 03:11:37
●	0.61	174.62.207.20	471.97		0.23	0.00	2012-08-05 03:11:39
	0.6	209.6.232.79	47.44		0.00	0.00	2012-04-12 19:41:07
●	0.63	76.175.169.116	773.54		10.30	0.00	2012-08-05 03:11:34
●	0.63	24.34.221.134	1434.77		0.80	0.00	2012-08-05 03:11:39
	0.63	24.147.69.225	4523.30		2086.77	0.00	2012-05-27 09:24:04
●	0.63	75.67.17.113	777.22		0.47	0.00	2012-08-05 03:11:42
	0.6	24.218.216.22	0.24		0.00	0.00	2012-02-26 16:12:48

Open Infrastructure Deployment

- Since Feb 2011:
  - 30 home APs: Boston and SF Bay
  - 1.3TB data trace over 6 months
  - 115 million network usage records and counting

- Spans 2 major ISPs
  - Comcast
  - RCN

Leveraging Residential Devices

- Can residential installations provide these services?

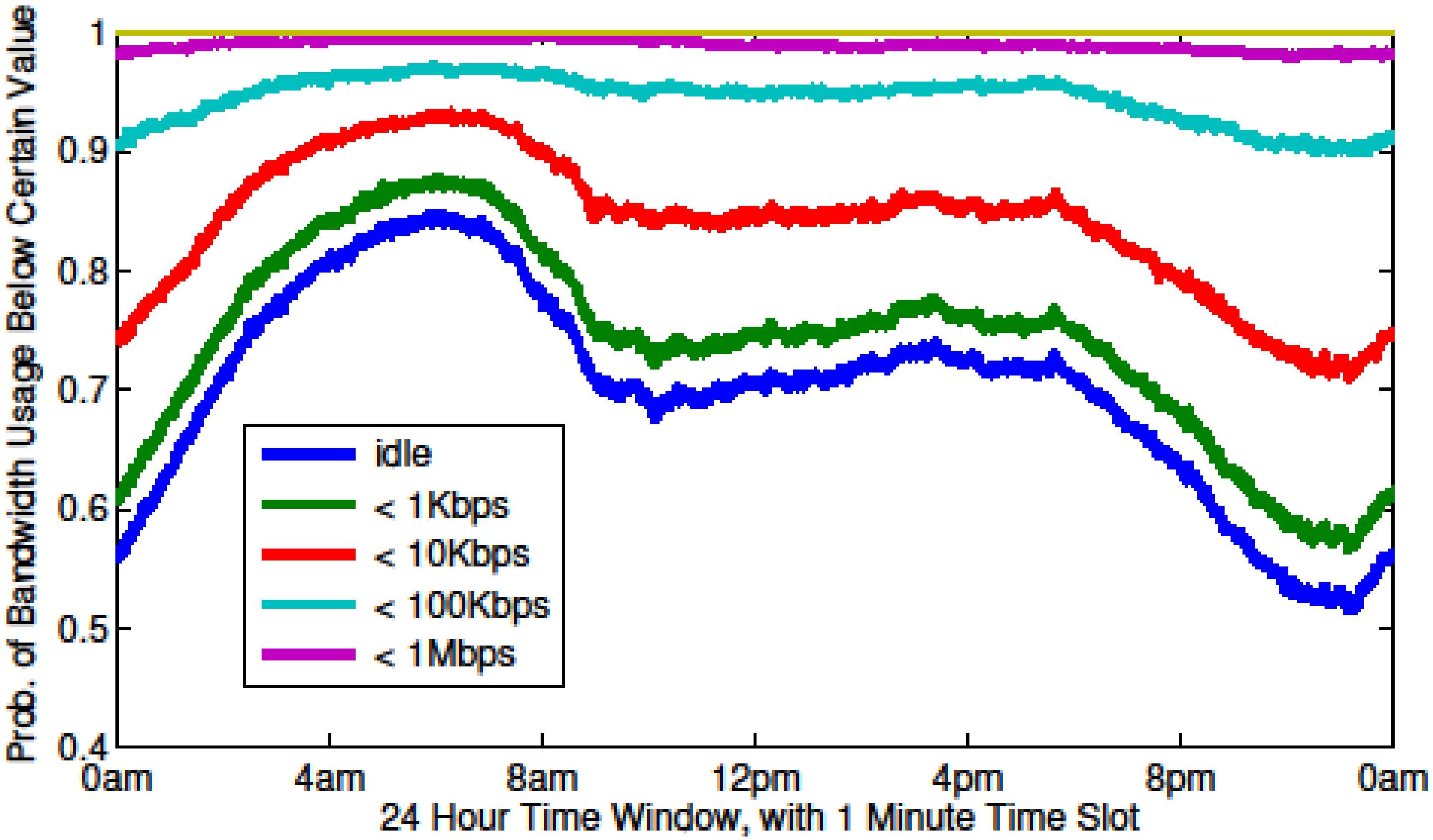
- Network Access Coverage
  - How dense is urban AP deployment?
    - Boston: 17 average, 7 reachable [JinTao2013]

- Cloud Services
  - Is there enough uplink to share?
  - How much latency can be expected?
  - How will services impact home traffic?

- **Used OpenInfrastructure to provide answers**

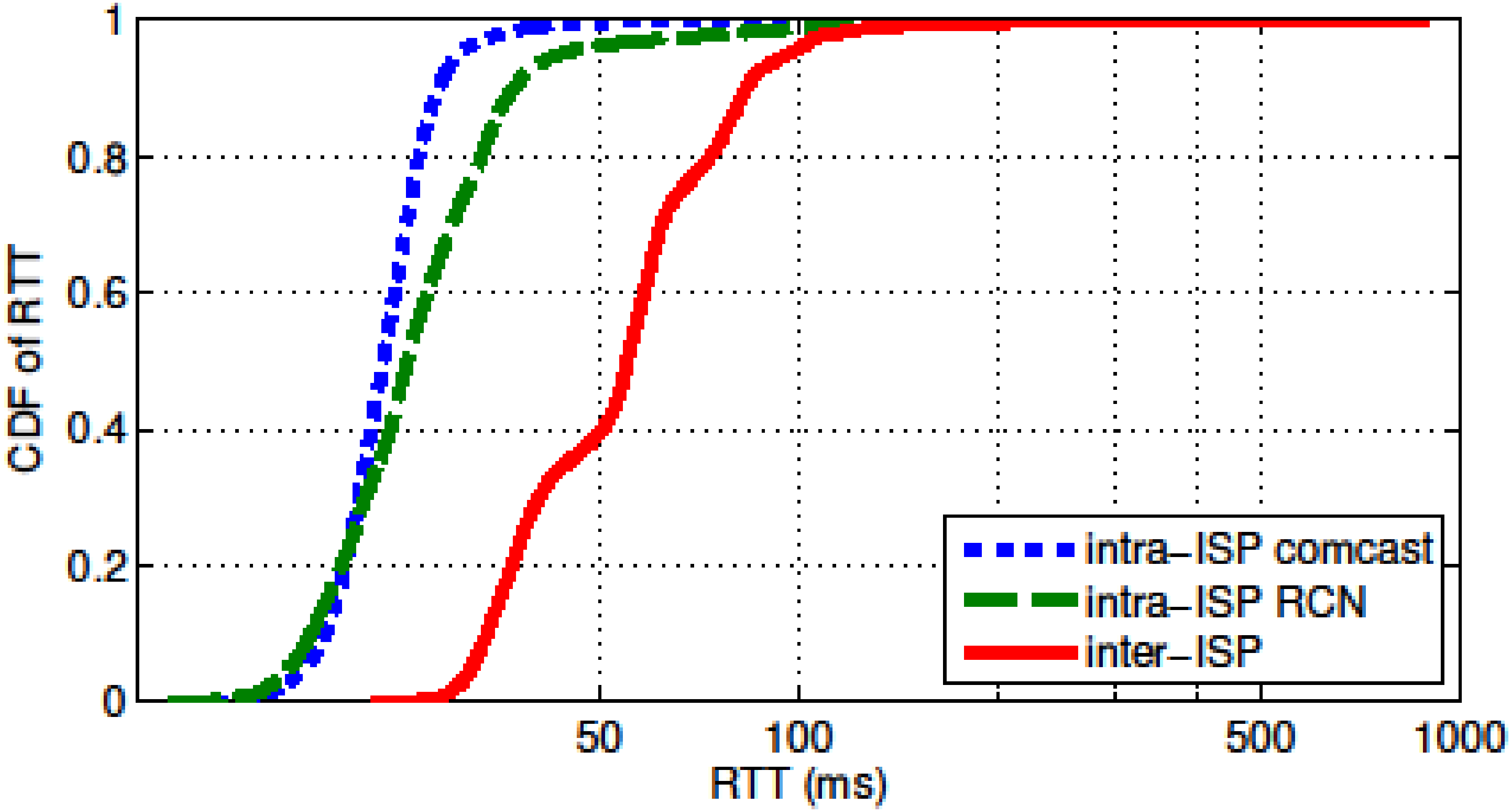
Residential Backhaul Usage Patterns

- Deployment data trace uplink: backhaul is underutilized
  - Results consistent with related, more limited work [Marcon2011]



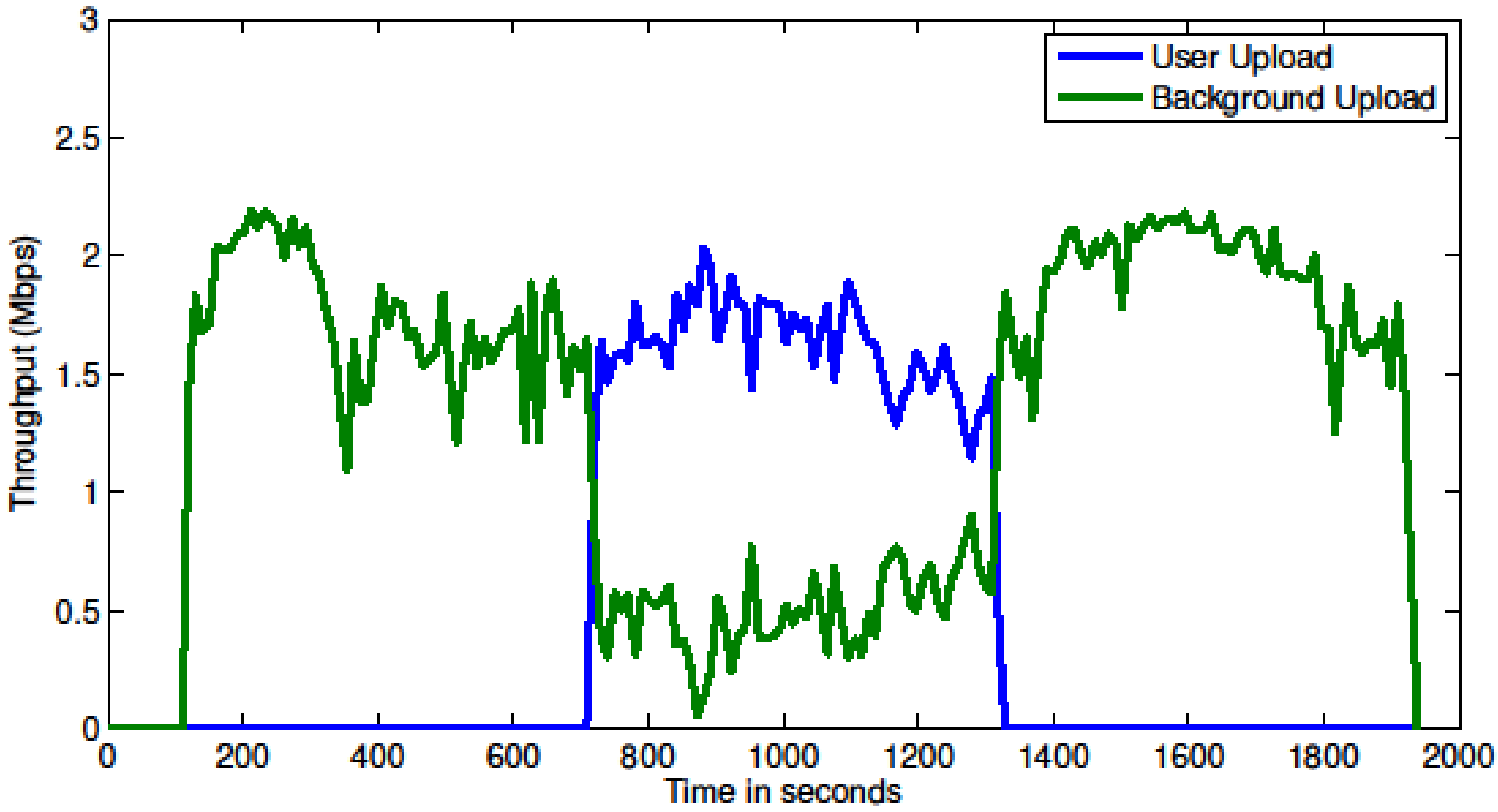
Testbed RTT

- RTT within OpenInfrastructure and CDNs



Background Throughput Impact

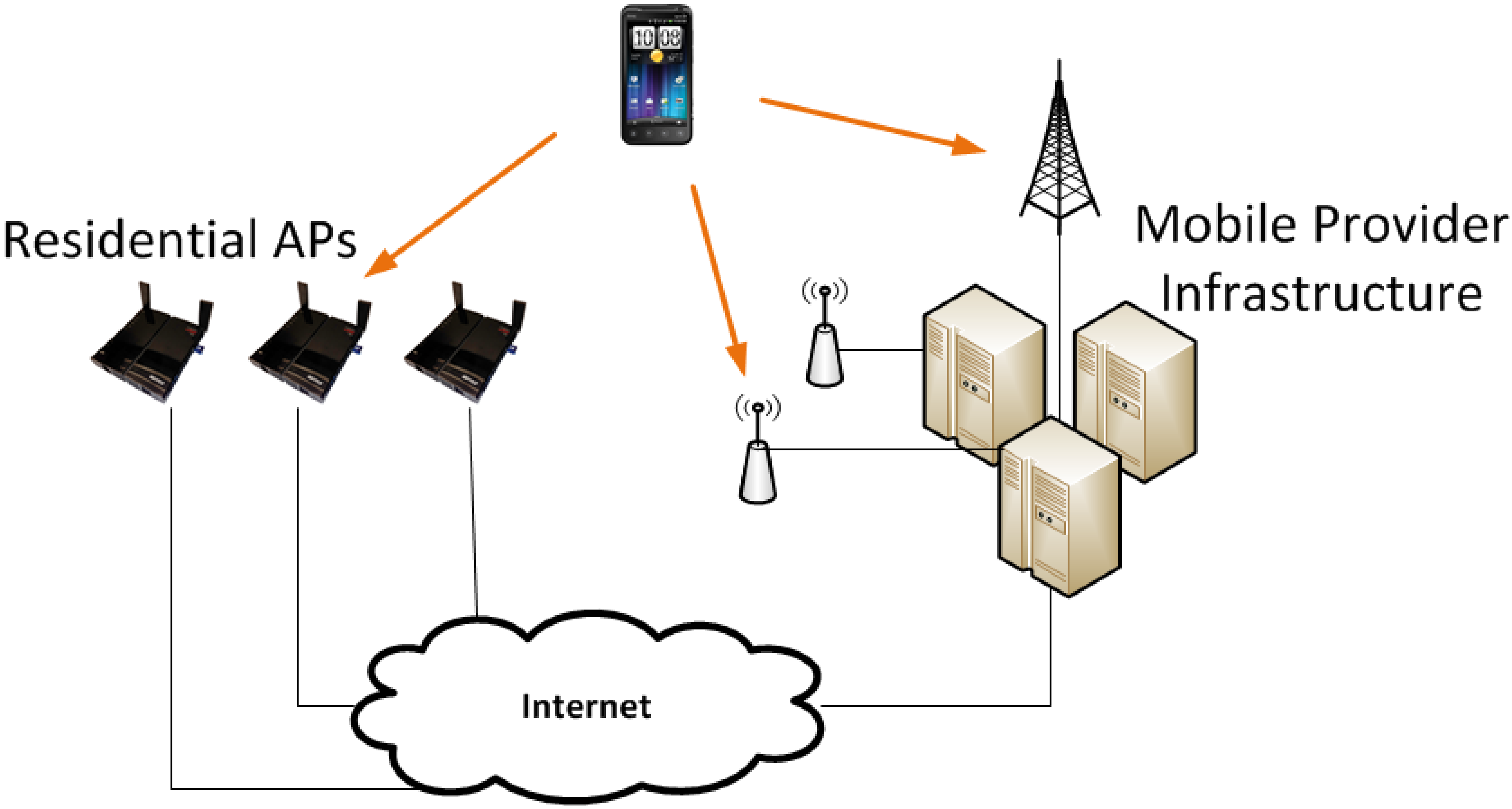
- Concurrent uplink usage test



Overview

1. Open Infrastructure
2. Residential Network Access
3. Edge Storage
4. Schedule
5. Questions

Providing Network Access



Wi-Fi Access Control Today

- Wi-Fi offloading from carriers is substantial (30% of total) [CISCO2013]

- 4G Standards include offloading mechanisms [3GPP TS 23.261]

- Options for access control:
  - WPA and EAP mechanisms allow confidentiality and control
  - WPA-Enterprise – uses username/passwords over tunnel
  - WPA-SIM – uses SIM card in handset
  - Open + Captive Portal

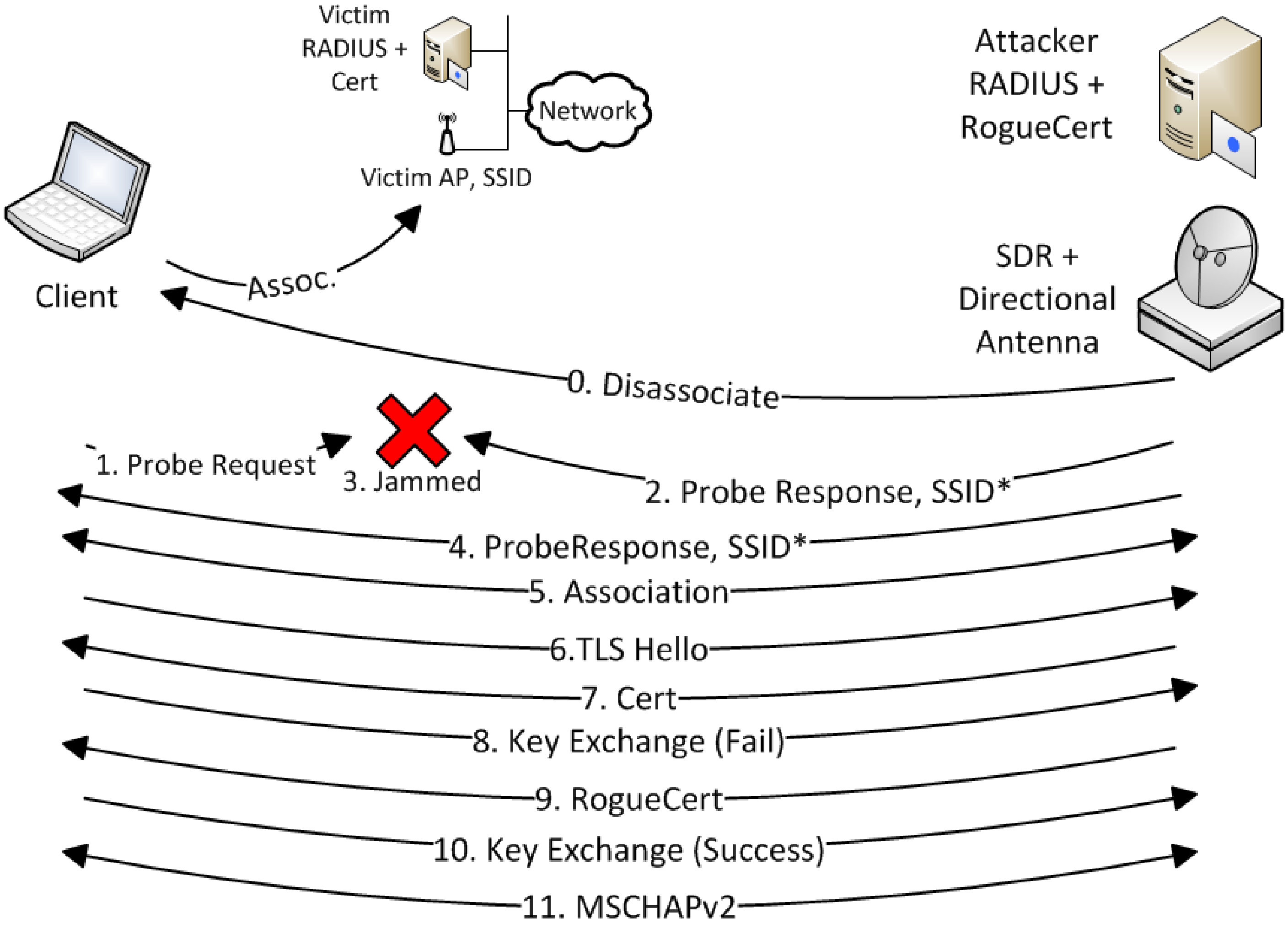
Risks in Wi-Fi

- Wi-Fi systems vulnerable to impersonation (Evil Twins)
  - [Damsgaard2006], [Bauer2008], [Gonzales2010]

- WEP, WPA key derivation
  - WEP [Bittau2006]
  - TKIP [Tews2009]
  - WPA Cracking [Marlinspike2012]

- New attacks can exploit multilayer weaknesses to steal credentials [Cassola2013]
  - Jamming prevents other APs on the set to reach client
  - Show new network identity, visually indistinguishable from original
  - Abuse password dialogs to hide creation of new profile
  - MITM, credential exposure

Stealthy Multi-layer Evil Twin Attack



State of Current Solutions

- Wi-Fi hotspots are commonly Open: AT&T, Xfinity, airports, Facebook Wi-Fi, etc.
  - Protection and confidentiality not widely deployed
  - Even if used, vulnerable, identity is revealed, need specialized maintenance

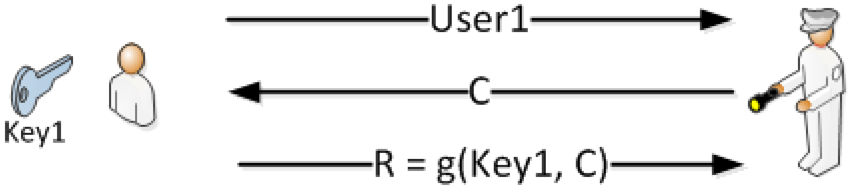
- Residential devices tie single network key to all identities
  - SSID key gives access to all who know the key
  - Second, public SSID and share key to all
    - Unique to device
    - Problem of key distribution
  - Revoking access is hard
  - Same service to all

Goals

- **Anonymous Authentication**
  - Provider gives access to a set of users S={U_1, U_2, … U_n}
  - U_i proves membership to the set without revealing its identity
- **Geographic untraceability**
  - Protect client- and AP owner’s IP from sites clients access
- **Low-overhead discovery**
  - Convenient client and provider signup
  - Identity establishment or agreement
- **Fine-grained access control**
  - Each set in S has a set of access limitations, enforced at AP
  - Incentive mechanisms

Anonymous Authentication

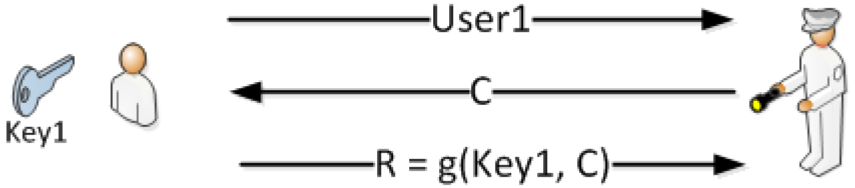
Authentication



User1	f(Key1)	Permission1
User2	f(Key2)	Permission2
User3	f(Key3)	Permission3
...		

Anonymous Authentication

Authentication


Key1

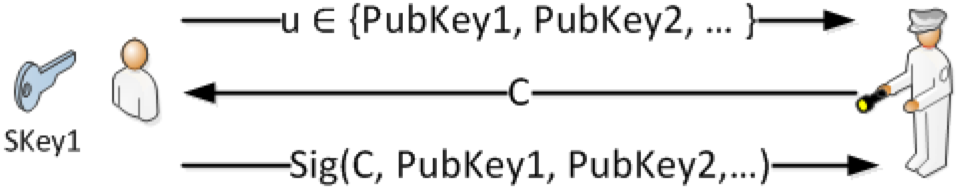
User1

C

R = g(Key1, C)

User1	f(Key1)	Permission1
User2	f(Key2)	Permission2
User3	f(Key3)	Permission3
…		

Anonymous Authentication


SKey1

$u \in \{\mathrm{PubKey1}, \mathrm{PubKey2}, \ldots\}$

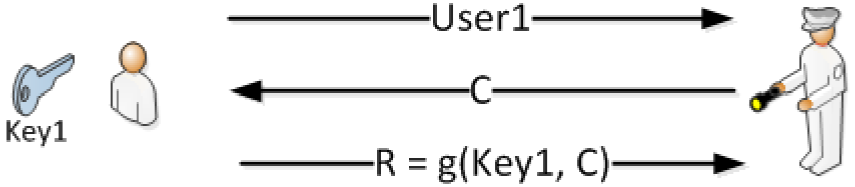
C

Sig(C, PubKey1, PubKey2,…)

User1	PubKey1	Permission1
User2	PubKey2	Permission2
User3	PubKey3	Permission3
…		

Anonymous Authentication

Authentication



User1 →

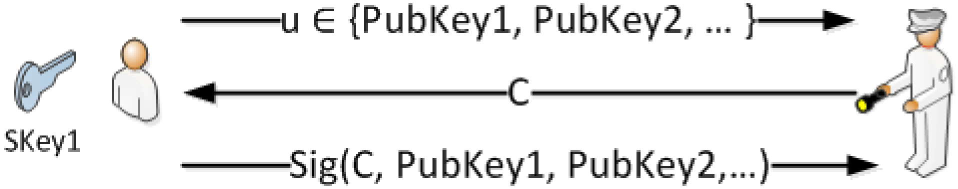
← C

R = g(Key1, C) →

Key1

User1	f(Key1)	Permission1
User2	f(Key2)	Permission2
User3	f(Key3)	Permission3
…		

Anonymous Authentication



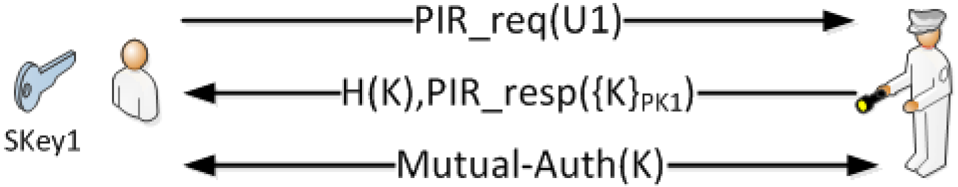
u ∈ {PubKey1, PubKey2, … } →

← C

Sig(C, PubKey1, PubKey2,…) →

SKey1

User1	PubKey1	Permission1
User2	PubKey2	Permission2
User3	PubKey3	Permission3
…		



PIR_req(U1) →

← H(K),PIR_resp($\{K\}_{PK1}$)

↔ Mutual-Auth(K)

SKey1

Group g, Key K

User1	PubKey1	$\{K\}_{PK1}$
User2	PubKey2	$\{K\}_{PK2}$
User3	PubKey3	$\{K\}_{PK3}$
…		

Anonymous Authentication

- Group signatures [Chaum91]
    - Supervising entity to reveal identities in case of dispute
    - Linear in size of anonymity set
- Ring Signatures [Rivest2001]
    - No supervisor
    - Also linear in |S|
- Computational Private Information Retrieval
    - First [Kushilevitz97]
    - Amortized O(log^2 n) comm. complexity [Gentry2005]
    - O(n/log n) pubkey ops [Lipmaa2009]

Fine-Grained Access Control

- Anonymity-only is easy to obtain: WPA-PSK
  - **Not flexible**

- Residential users may not wish to unrestricted access to all
  - Different service levels for users
  - Still maintain anonymity

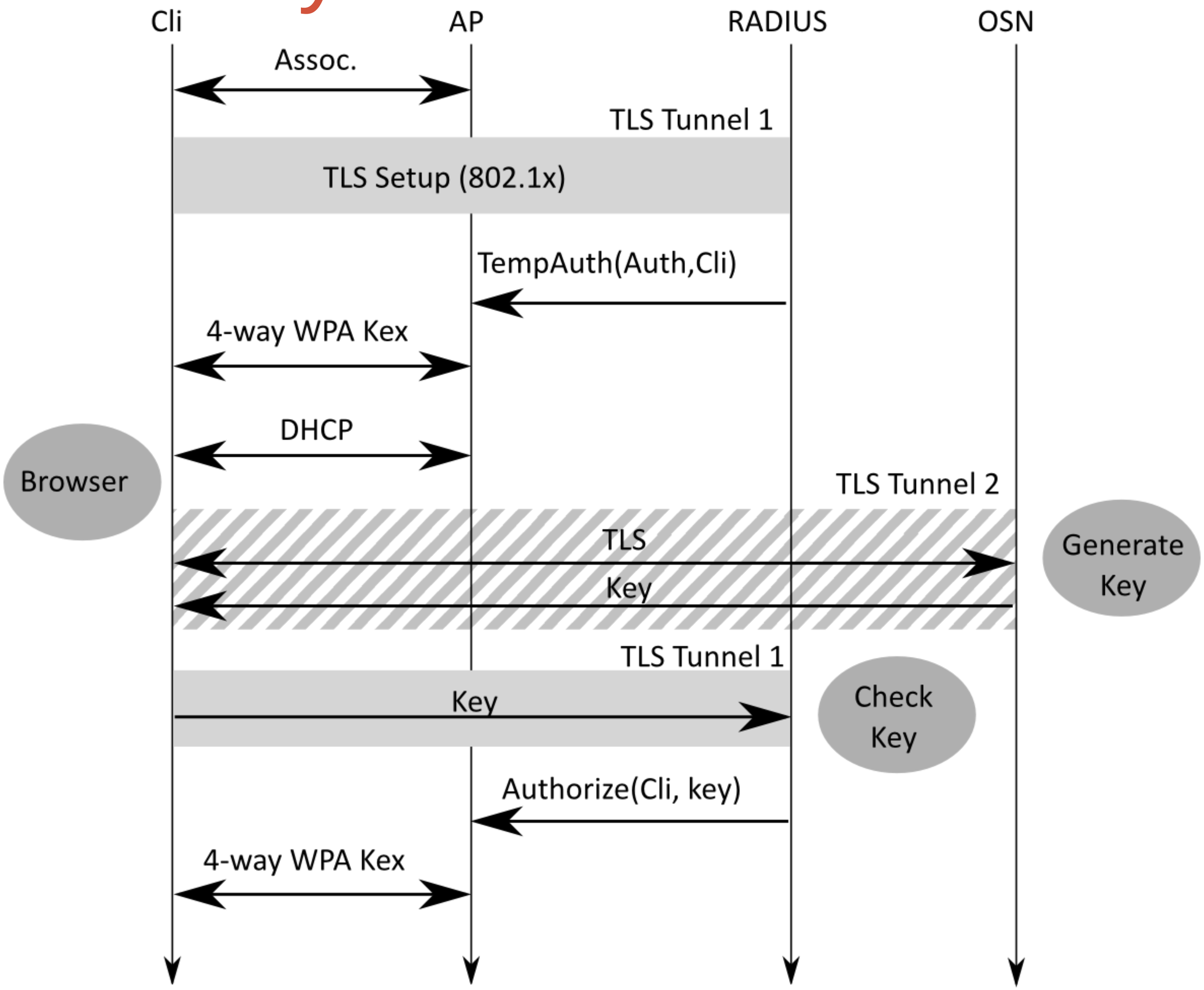
- Dynamic membership
  - Service may be terminated
  - New users may enter the set of served users

Low-Overhead Discovery Mechanism

- Users and providers need to meet before service is used
  - Establish identity
  - Exchange keys
  - Negotiate terms of use (payment, exchange, incentives)

- Leverage information in Online Social Networks
  - Public information as a directory of people and contact information (think PGP)
  - Still potential for impersonation

Preliminary work: SNEAP



Features, Limitations and Future Work

- SNEAP Features:
  - Solves the SSID-Certificate problem
  - Uses OSN API features to decide link between user/AP
  - Provides encrypted link early
  - ➢Facebook-Cisco’s Wi-Fi is plaintext
- Limitations
  - User and AP owner identities are revealed to each other when connecting
  - OSN knows User-Provider link
- Future Work
  - **Anonymous authentication method, Sybill protection, perfomance**
  - OSN as directory
  - Incentives

Overview

1. Open Infrastructure
2. Residential Network Access
3. Edge Storage
4. Schedule
5. Questions

Cloud Storage Today

- Large providers (GDrive, Dropbox, Microsoft, Wuala, etc)
  - Heterogeneous privacy protection
  - Centrally managed storage (own infrastructure)
  - Delegated storage (S3, Azure)

- Personal Cloud / File sharing (owncloud, BTSync, WD MyCloud)
  - Storage is user-hosted
  - Mostly single user / some hosting capabilities (owncloud)
  - Some privacy

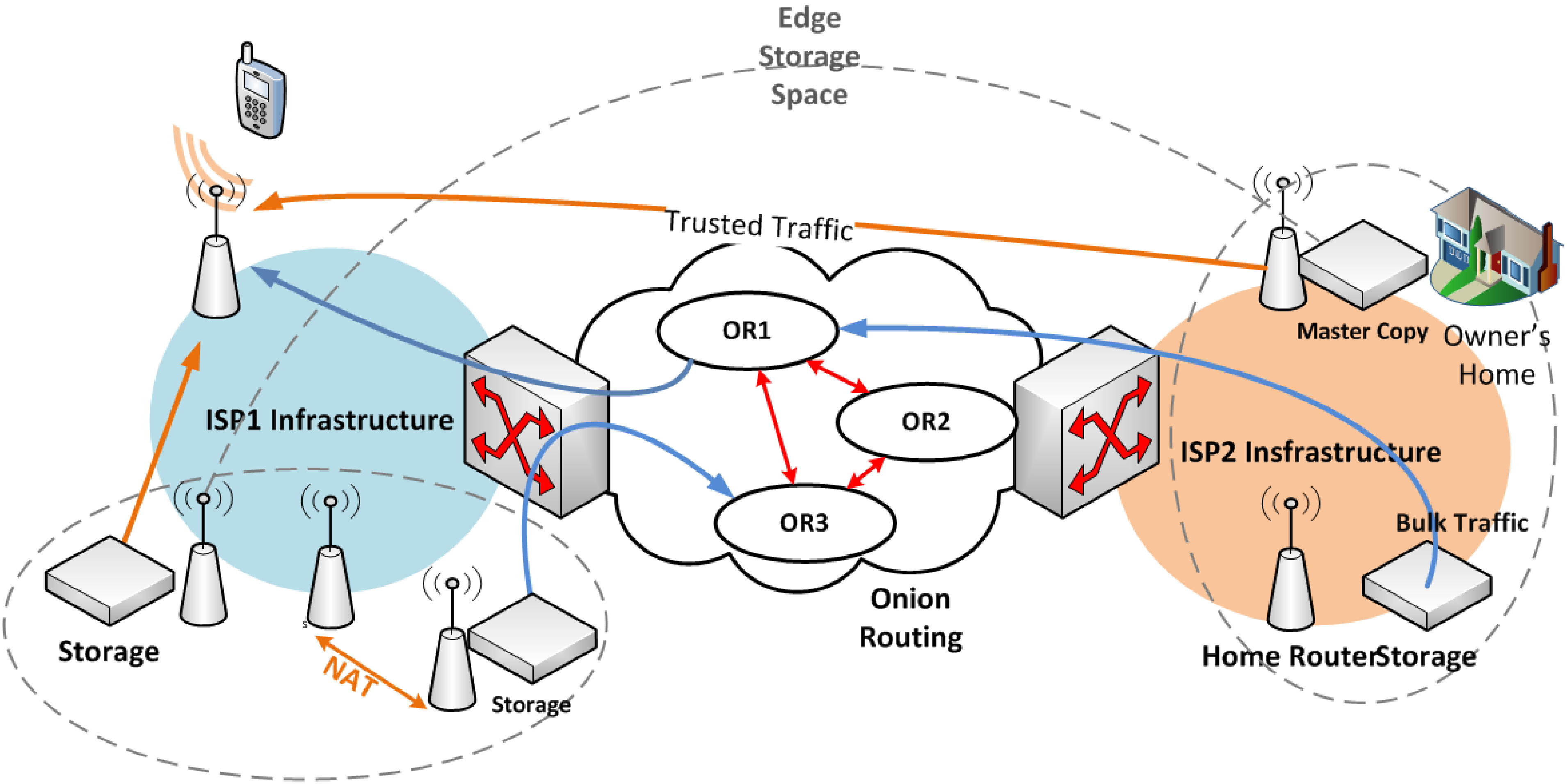
Privacy Pitfalls

- Clients access services directly, exposing IP
- IP Anonymizing (TOR) is not straightforward
  - No support for UDP communications
  - Side-channel leaks (DNS queries)
- Service + EncFS/Truecrypt + TOR
  - User identity revealed to service provider through authentication
  - Client program can leak or reveal information
    - Local daemon can read IP and already monitors FS activity
  - Access patterns

SafEdge Storage Services

- **Goal: Private and efficient anonymous storage**
- Performance
  - **High throughput, low-impact**
  - **Low overhead**
  - Incentive mechanisms
- Untraceability: session endpoint hiding
- Content Protection
  - Transport encryption
  - Data confidentiality
  - Resiliency
  - Access Pattern Protection

SafEdge Storage Architecture

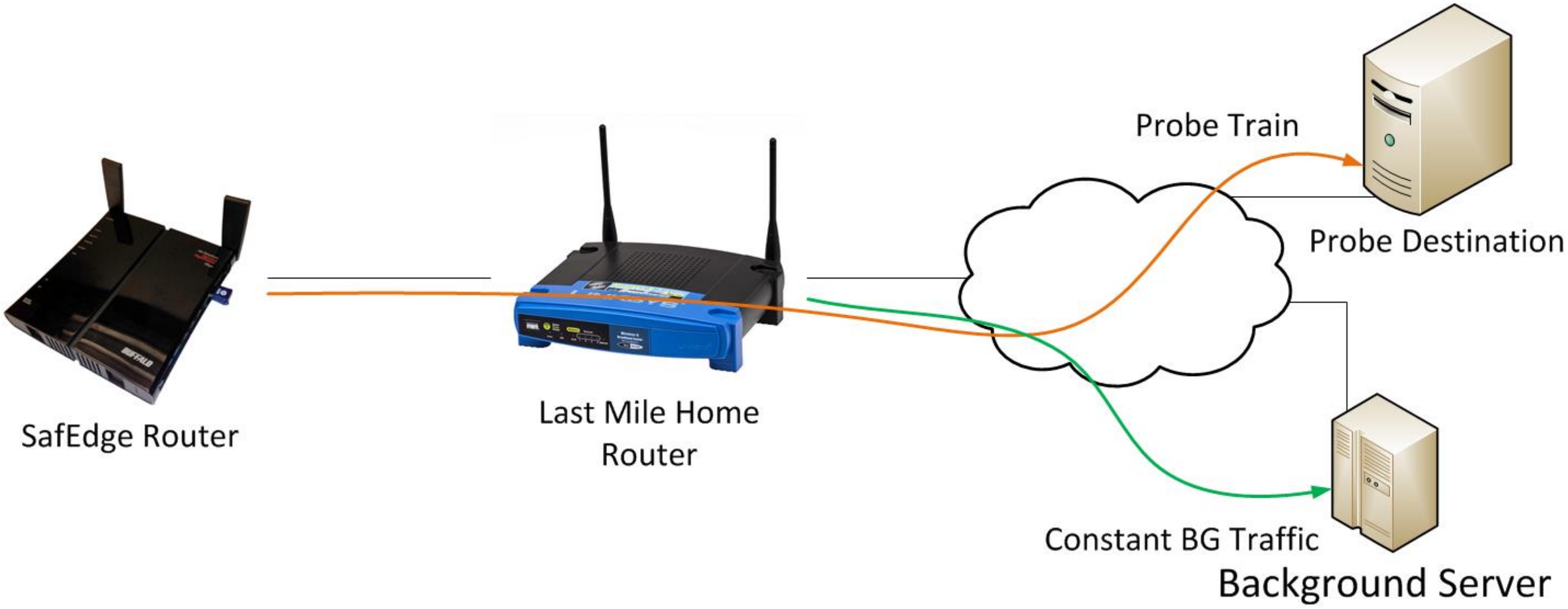


SafEdge Throughput

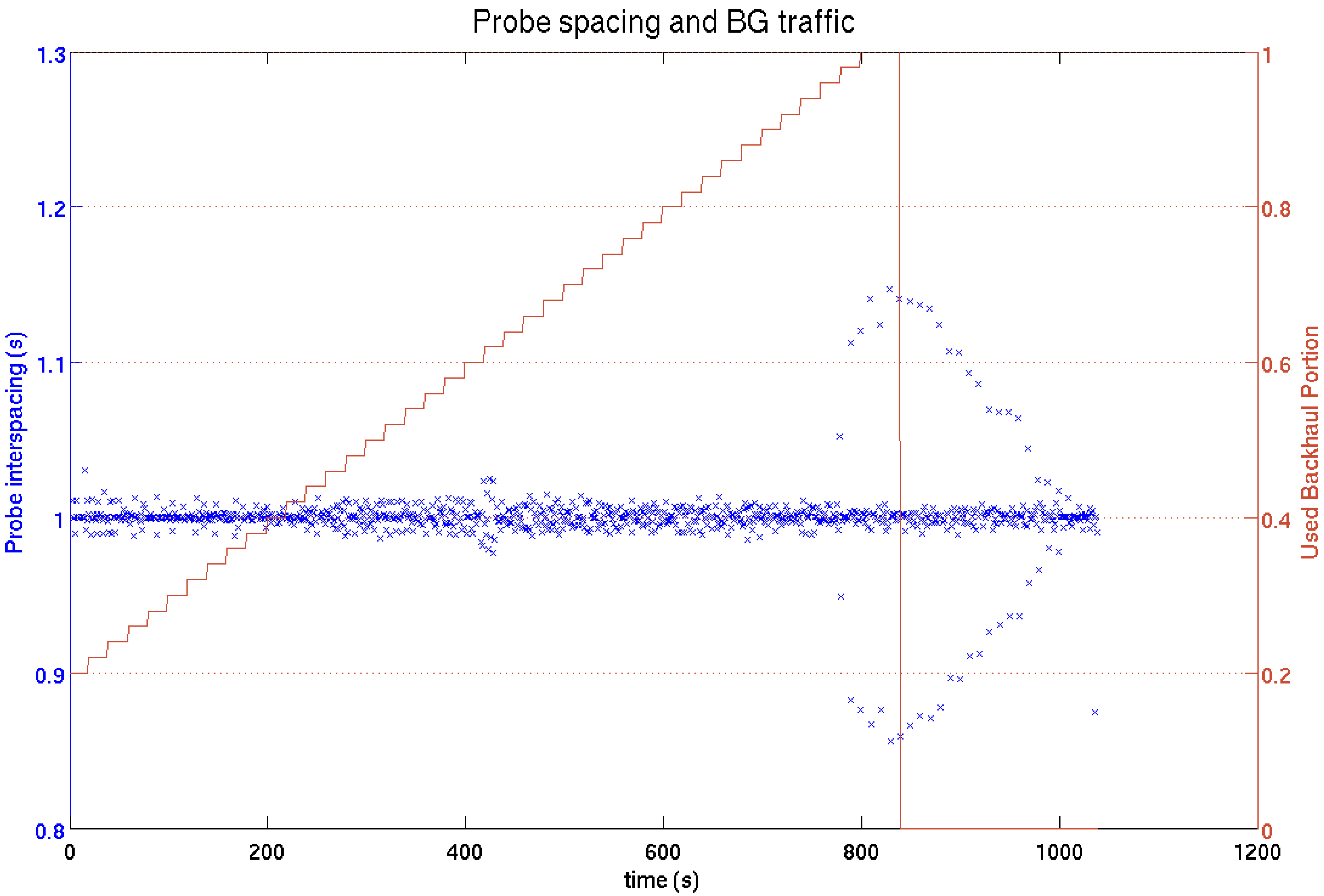
- SafEdge Storage runs on uplink-limited residential links
  - ➢Prioritize regular home traffic
- Two scenarios
  - Component runs with full view of last mile link.
  - Component runs behind another device (typically NAT)
    - ➢Application must back-off when gateway saturated

- Onion routing can be slow
  - TCP throughput over TOR is limited by node owners
  - Large latency
  - ➢Have Master Copy coordinate, client aggregates links

Characterizing Shared Throughput



Bandwidth Probing



Existing and Future work

- Client-provided cloud storage [Zhou2012] [Zhang2013]
- Performance
  - Speed over the Onion aggregating storage providers:
    - Throughput aggregation [Kandula2008] [Jin2013]
    - Performance of hidden services [Loesing2009] [Snader2009,2011]
  - Uplink congestion detection
    - Available Bandwidth [Jain2002]
  - ➢ Performance measurement over OpenInfrastructure
- Privacy Protection
  - Endpoint hiding, hidden services [TOR2004], ORAM [Stefanov2013]
  - Storage and transport confidentiality

Summary and Takeaway

- Cloud services and wireless network access as they stand today offer uneven privacy guarantees

- Edge services that leverage large numbers of participants can help mitigate privacy risks

- Research in this area brings about interesting services and research problems
  - Characterization of urban residential networks
  - Anonymous Wi-Fi authentication
  - Efficient, well-behaved Edge storage

Proposed Schedule

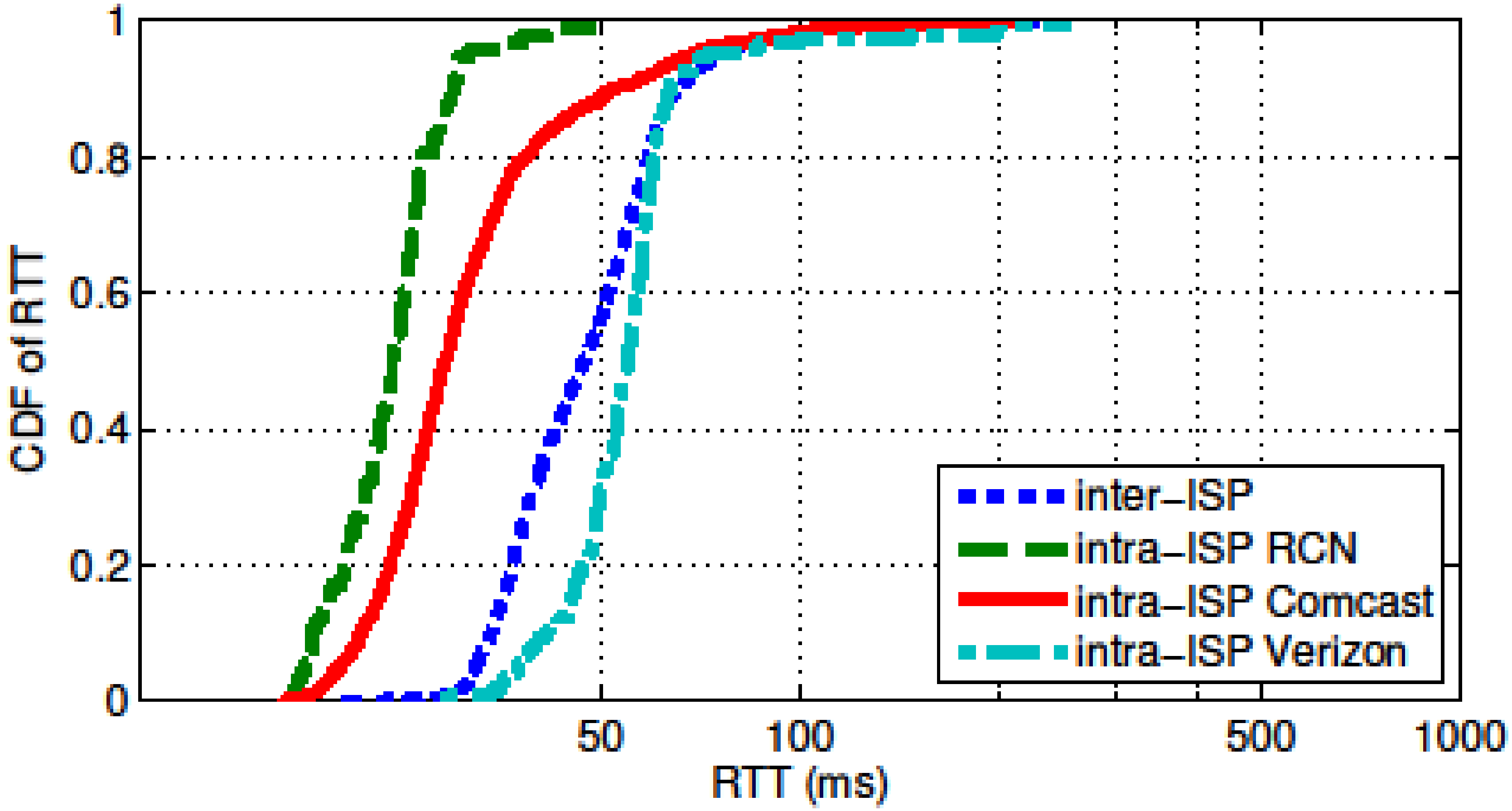
Proposed Task	Completion Date (by end of)
Anonymous Wi-Fi Authentication Design and Implementation	February 2014
Storage, Throughput Aggregation Design	March 2014
Storage and Throughput Implementation	April 2014
Performance Evaluation	May 2014
Dissertation defense	June 2014

Thank you!

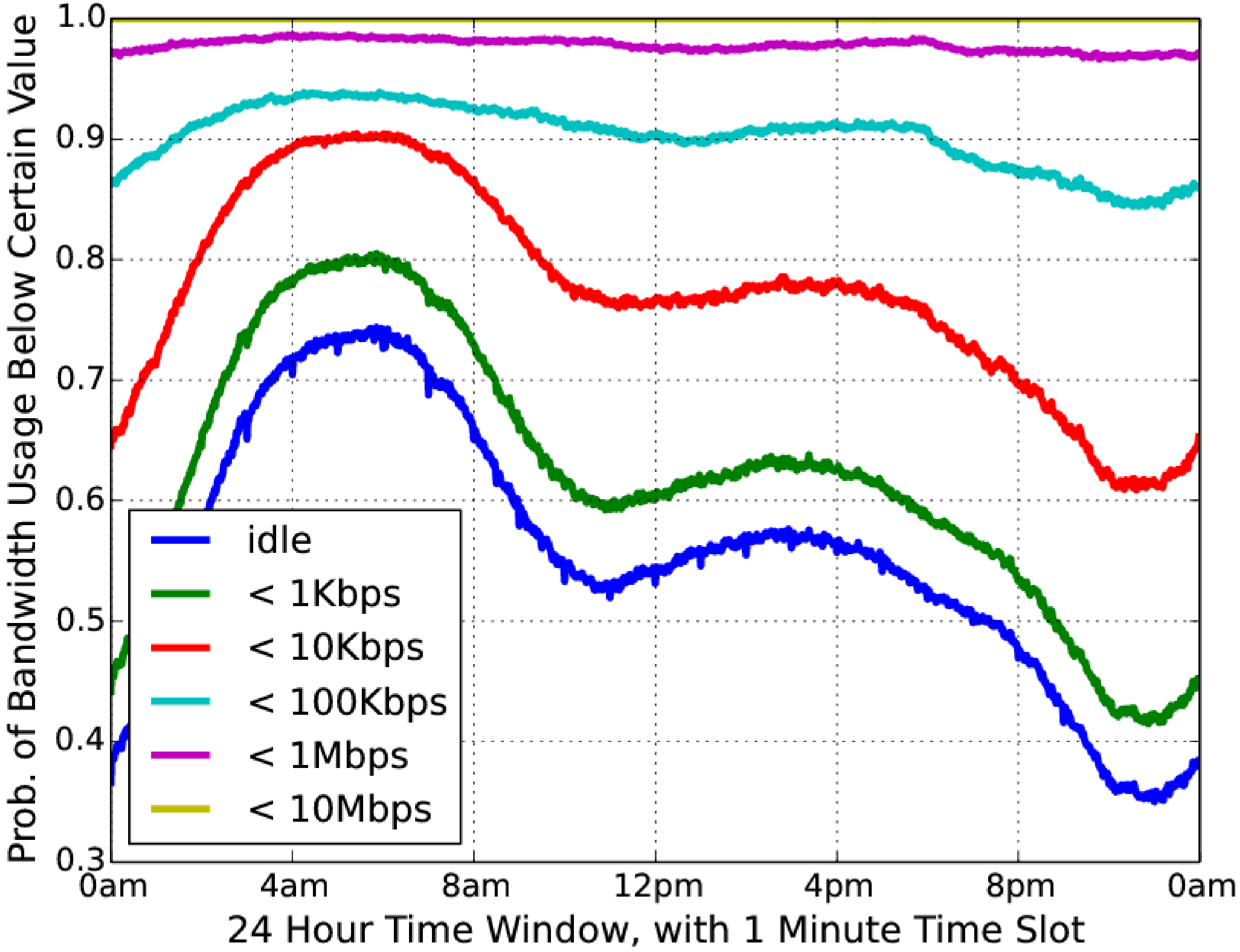
Q&A

Density and Residential Round-Trip Time

- Wardriving ping test (Urban Boston) [JinTao2013]
  - 17 visible APs at any time, 7 reachable on avg.



Bandwidth Usage (Nov ’12-Feb ’14)


Prob. of Bandwidth Usage Below Certain Value
0am
4am
8am
12pm
4pm
8pm
0am
24 Hour Time Window, with 1 Minute Time Slot
idle
< 1Kbps
< 10Kbps
< 100Kbps
< 1Mbps
< 10Mbps