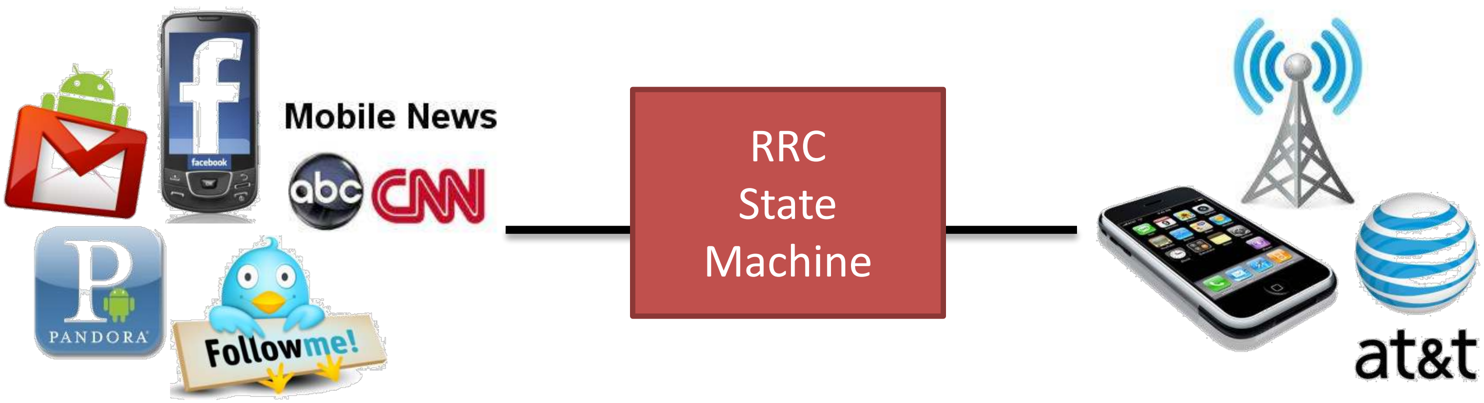
Profiling Resource Usage for Mobile Applications: *a Cross-layer Approach*

Feng Qian[1], Zhaoguang Wang[1], Alexandre Gerber[2], Z. Morley Mao[1], Subhabrata Sen[2], Oliver Spatscheck[2]

[1]University of Michigan [2]AT&T Labs - Research

Introduction

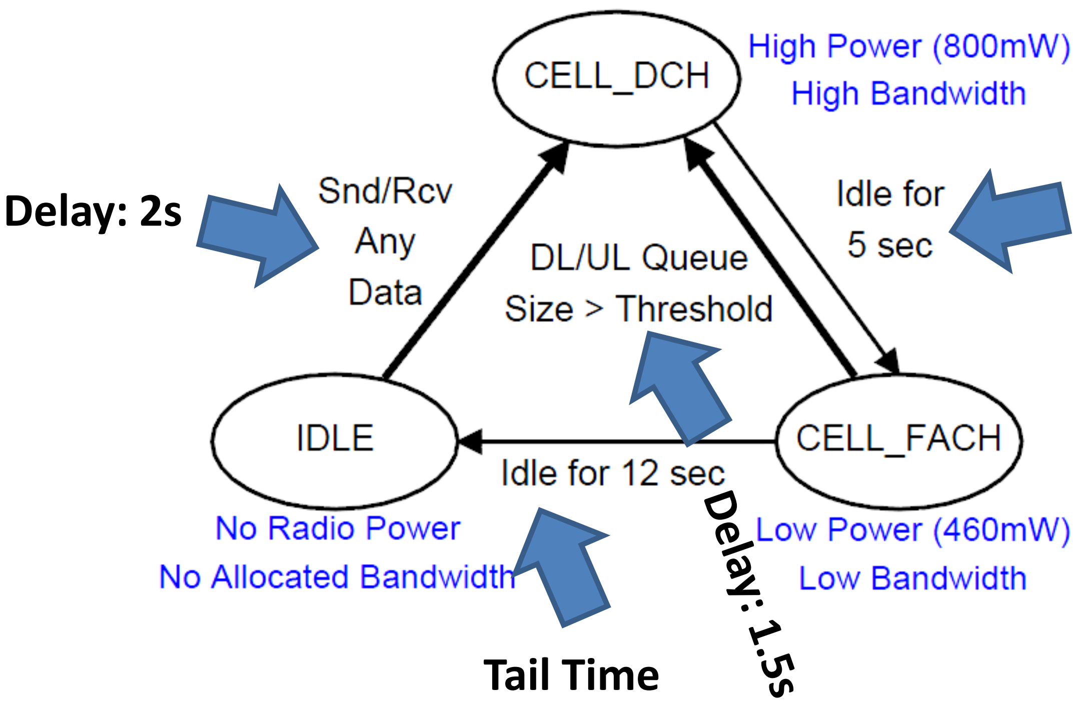
- Typical testing and optimization in cellular data network



- Little focus has been put on their **cross-layer interactions**
  Many mobile applications are not cellular-friendly.
- The key coupling factor: the **RRC State Machine**
  - Application traffic patterns trigger state transitions
  - State transitions control radio resource utilization, end-user experience and device energy consumption (battery life)

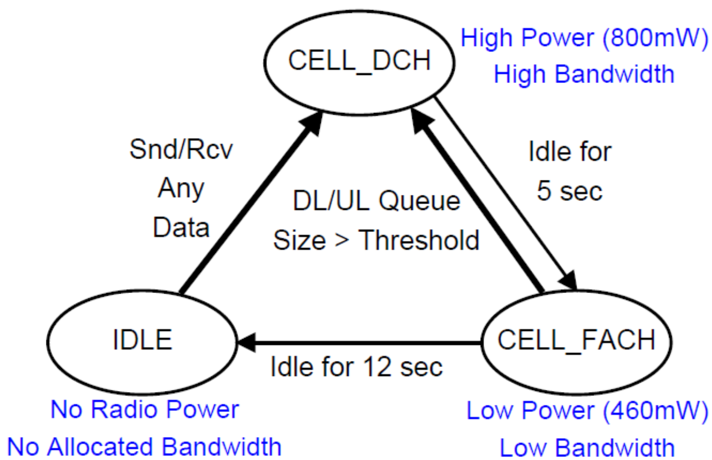
The RRC State Machine for UMTS Network

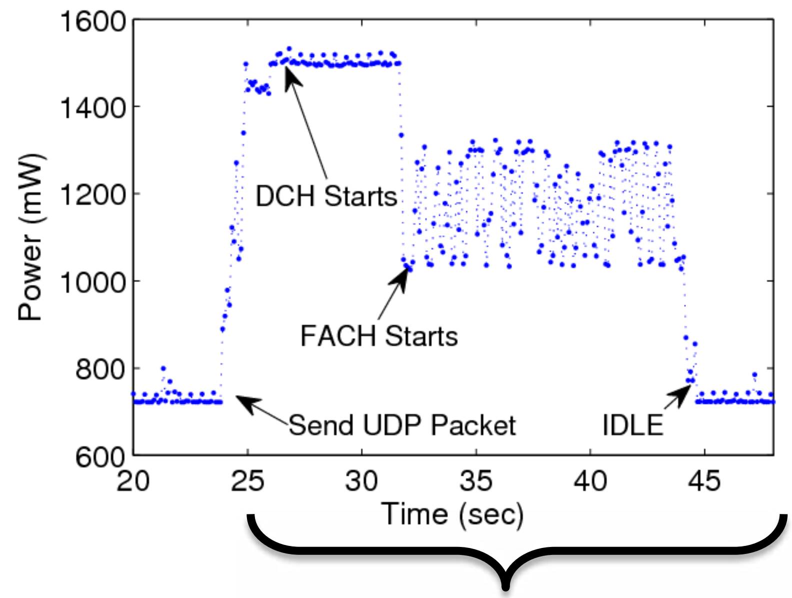
- State promotions have **promotion delay**
- State demotions incur **tail times**



	Channel	Radio Power
IDLE	Not allocated	Almost zero
CELL_FACH	Shared, Low Speed	Low
CELL_DCH	Dedicated, High Speed	High

Example: RRC State Machine for a Large Commercial 3G Network





Tail Time

Waiting inactivity timers to expire

DCH: High Power State (high throughput and power consumption)

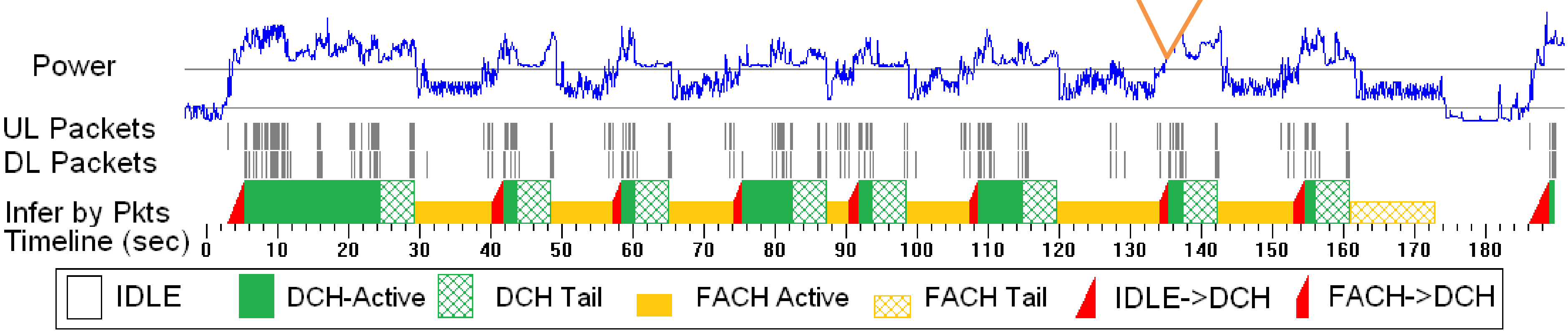
FACH: Low Power State (low throughput and power consumption)

IDLE: No radio resource allocated

Example of the State Machine Impact: Inefficient Resource Utilization

A significant amount of channel occupation time and battery life is wasted by **scattered bursts**.

State transitions impact end user experience and generate signaling load.



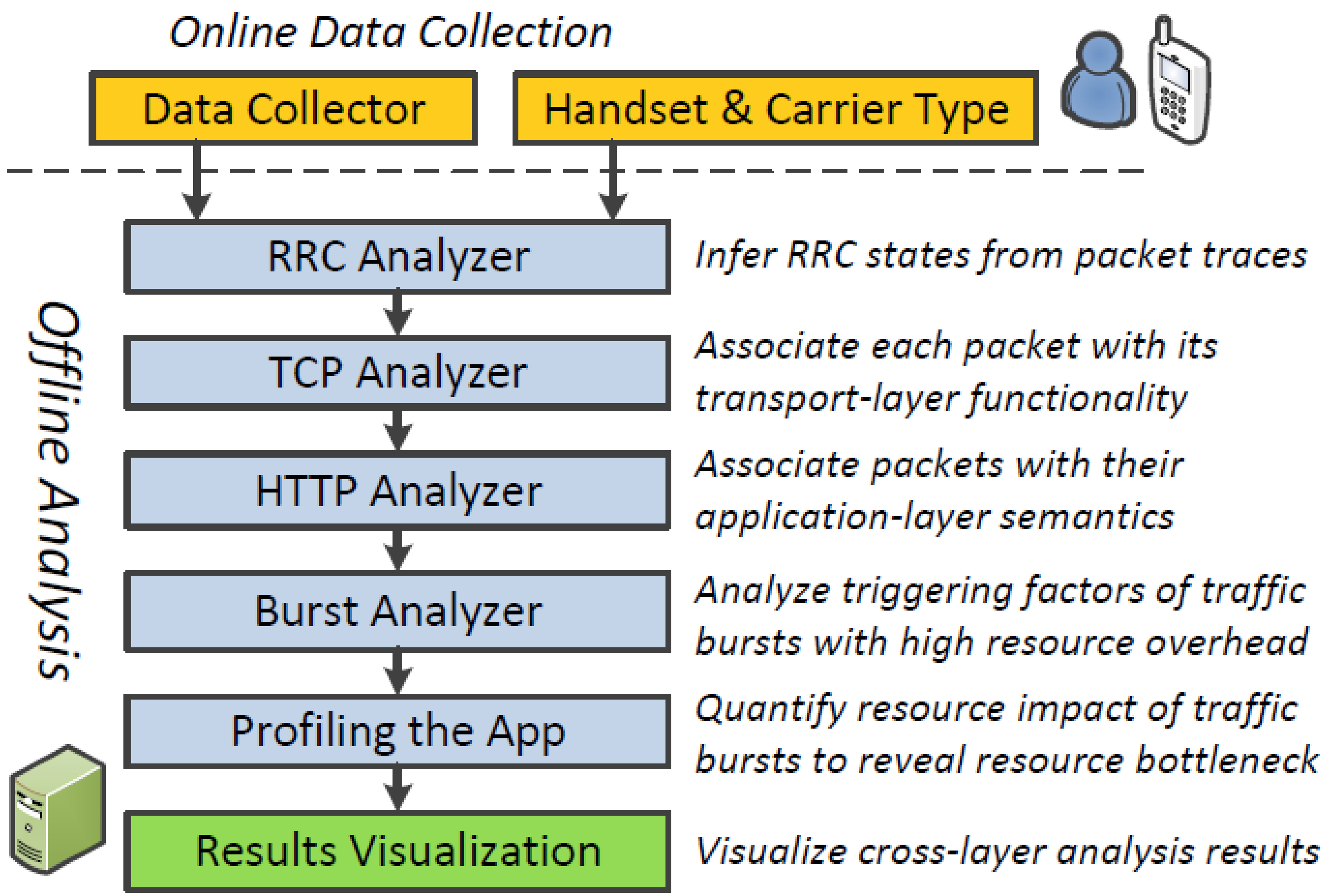
Analysis powered by the ARO tool

	FACH and DCH
Wasted Radio Energy	34%
Wasted Channel Occupation Time	33%

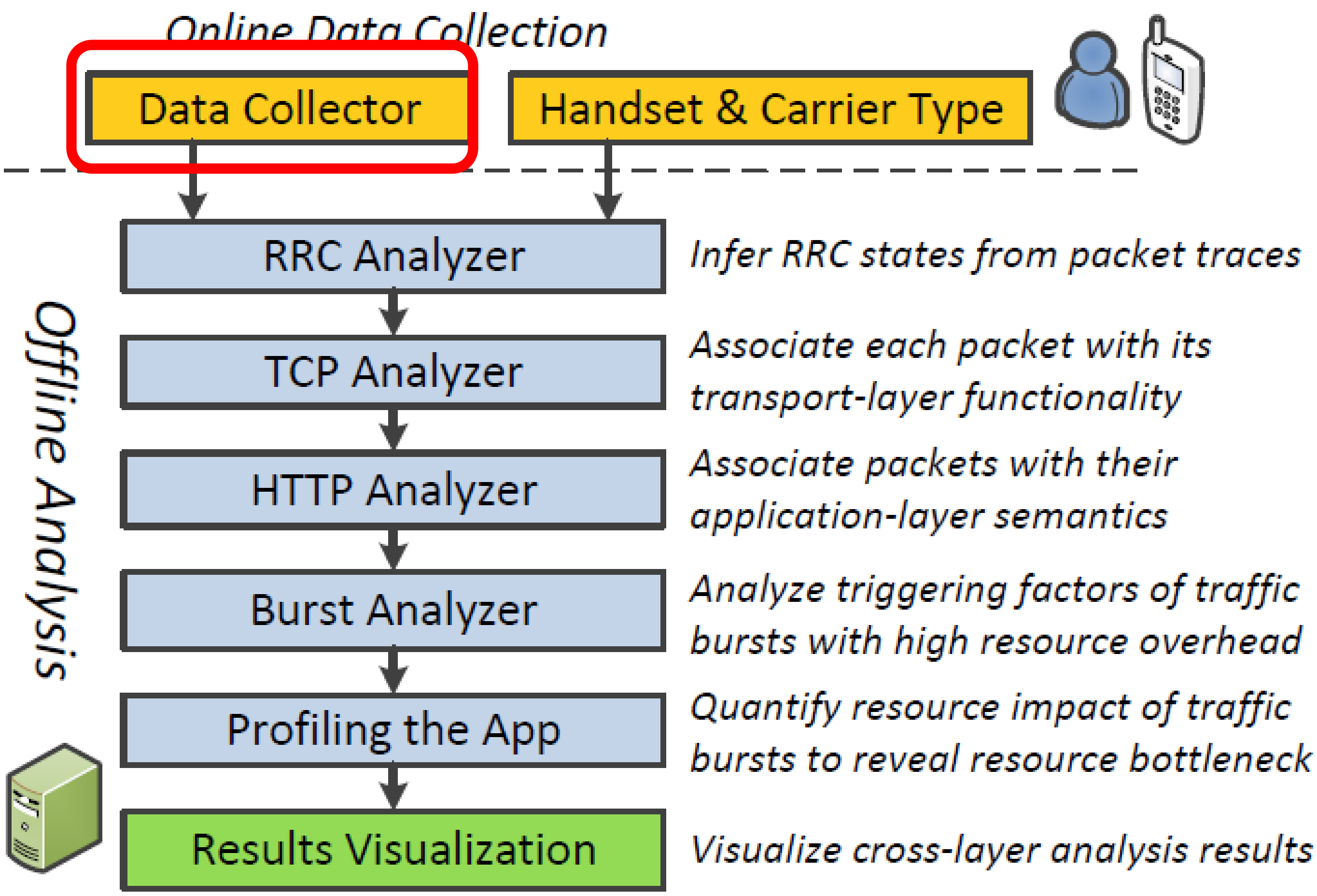
ARO: Mobile Application Resource Optimizer

- **Motivations:**
  - Are developers aware of the RRC state machine and its implications on radio resource / energy? **NO.**
  - Do they need a tool for automatically profiling their prototype applications? **YES.**
  - If we provide that visibility, would developers optimize their applications and reduce the network impact? **Hopefully YES.**
- **ARO: Mobile Application Resource Optimizer**
  - Provide visibility of radio resource and energy utilization.
  - Benchmark efficiencies of cellular radio resource and battery life for a specific application

ARO System Architecture



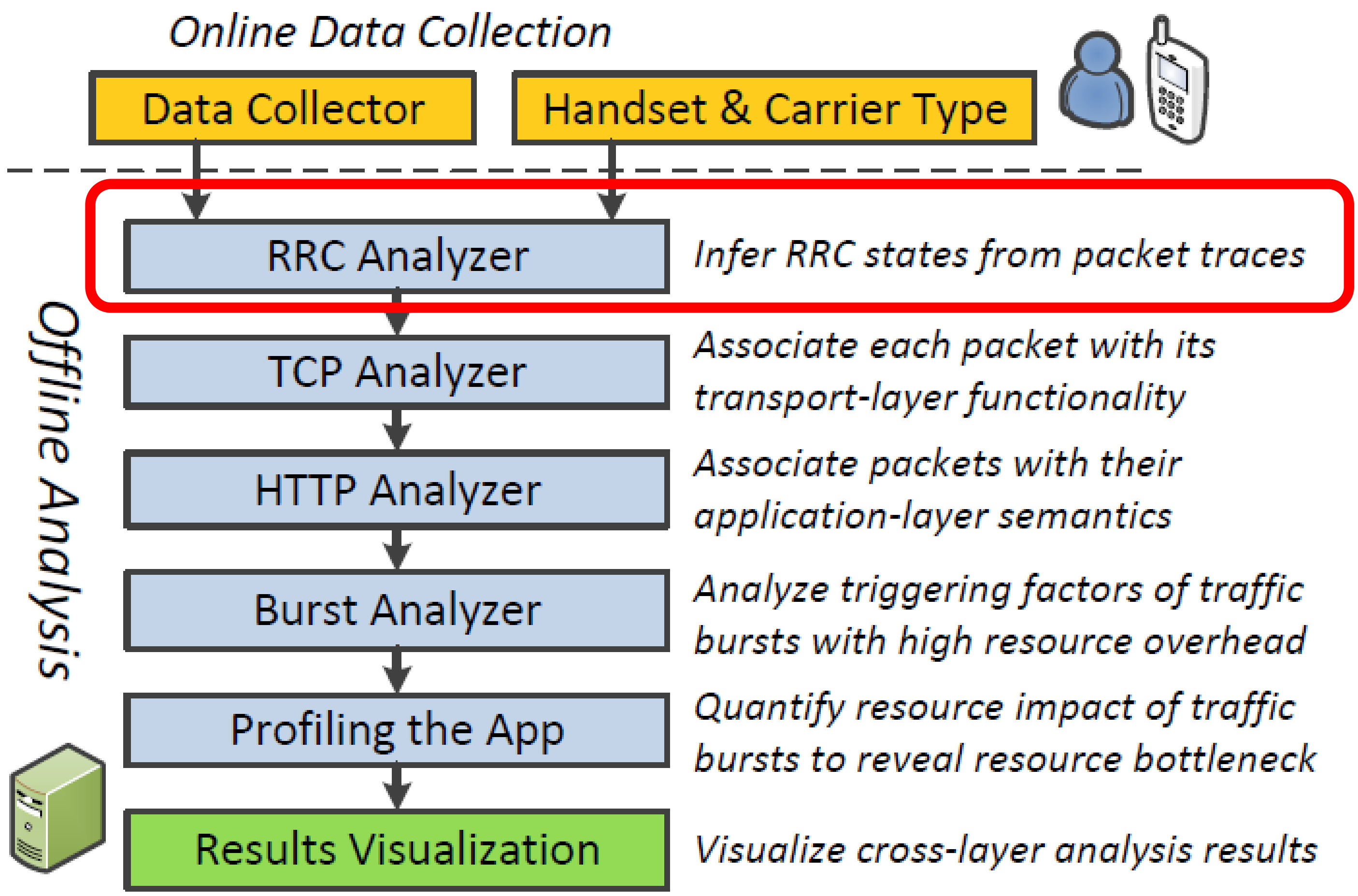
ARO System Architecture



The Data Collector

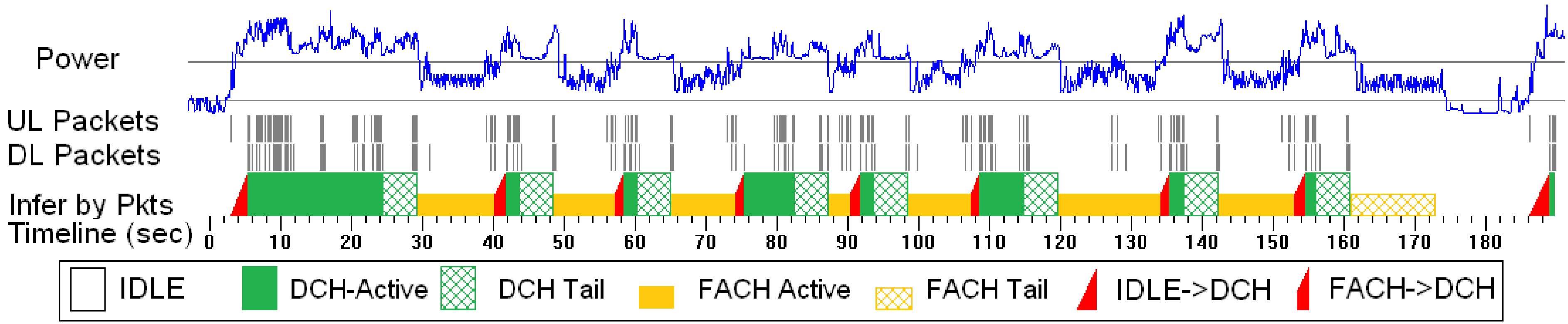
- Collects three pieces of information
  - The packet trace
  - User input (e.g., touching the screen)
  - Packet-process correspondence
    - The RRC state transition is triggered by the **aggregated** traffic of all concurrent applications
    - But we are only interested in our **target** application.
- Less than 15% runtime overhead when the throughput is as high as 600kbps

ARO System Architecture



RRC Analyzer: State Inference

- RRC state inference
  - Taking the packet trace as input, **simulate** the RRC state machine to infer the RRC states
  - Evaluated by measuring the device power



Example: Web Browsing Traffic on HTC TyTn II Smartphone

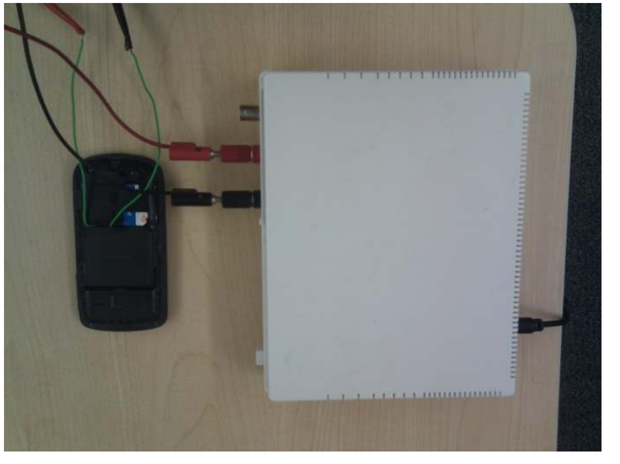
RRC Analyzer: Applying the Energy Model

- Apply the energy model
  - Associate each state with a constant power value
  - Based on our measurement using a power-meter

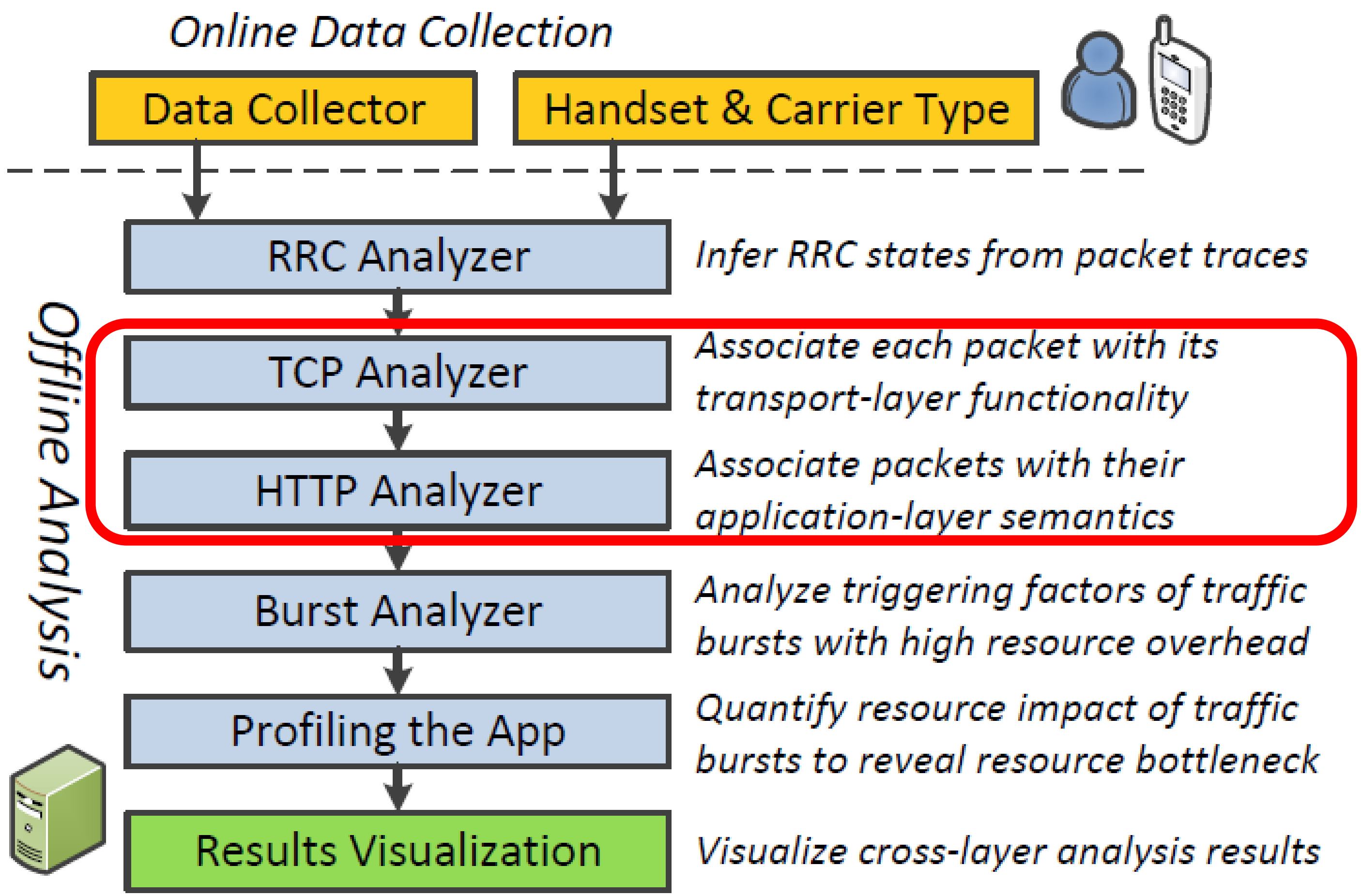
Table 3: Measured average radio power consumption

	TyTn Carrier 1	NexusOne Carrier 1	ADP1 * T-Mobile
$P(\text{IDLE})$	0	0	10mW
$P(\text{FACH})$	460mW	450mW	401mW
$P(\text{DCH})$	800mW	600mW	570mW
$P(\text{FACH}\rightarrow\text{DCH})$	700mW	550mW	N/A
$P(\text{IDLE}\rightarrow\text{DCH})$	550mW	530mW	N/A

* Reported by [27] for Android HTC Dream phone



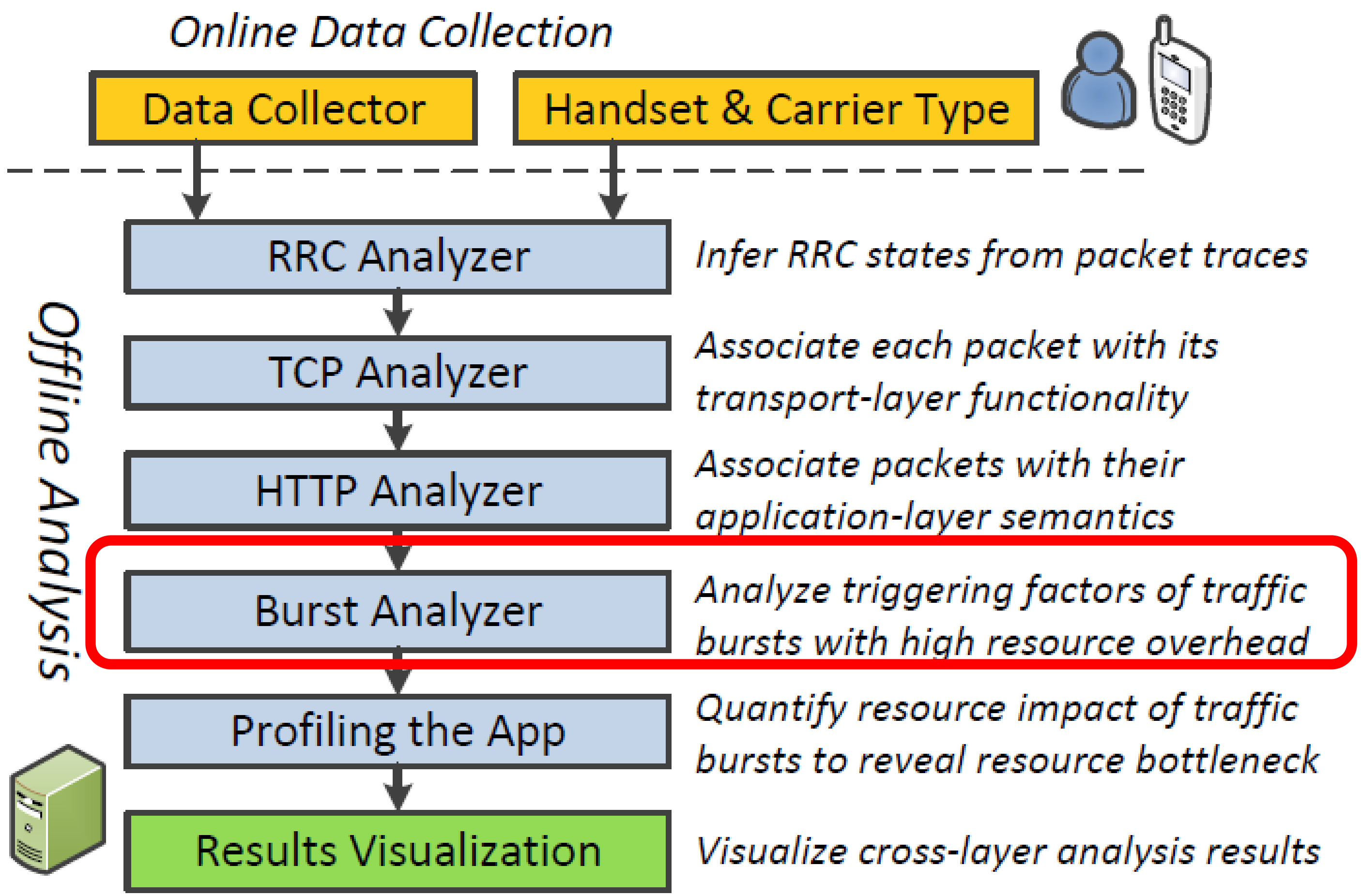
ARO System Architecture



TCP / HTTP Analysis

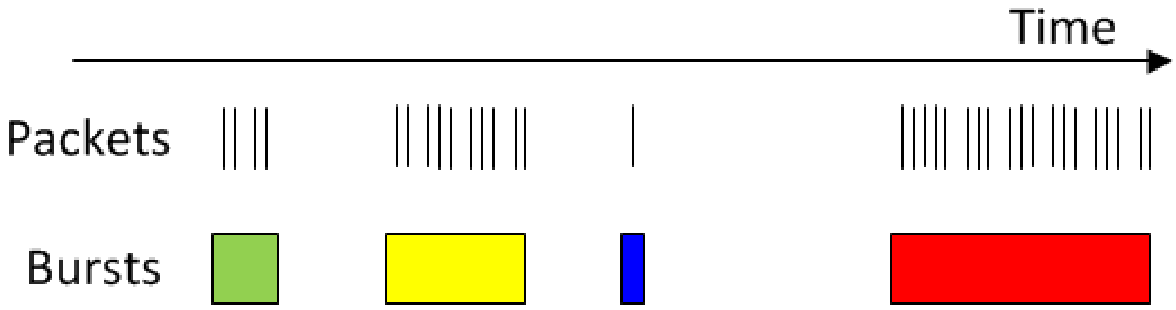
- TCP Analysis
  - Infer transport-layer properties for each TCP packet
    - SYN, FIN, or RESET?
    - Related to loss? (e.g., duplicated ACK / recovery ACK)
    - …
- HTTP Analysis:
  - HTTP is the dominant app-layer protocol for mobile apps.
  - Model HTTP behaviors

ARO System Architecture

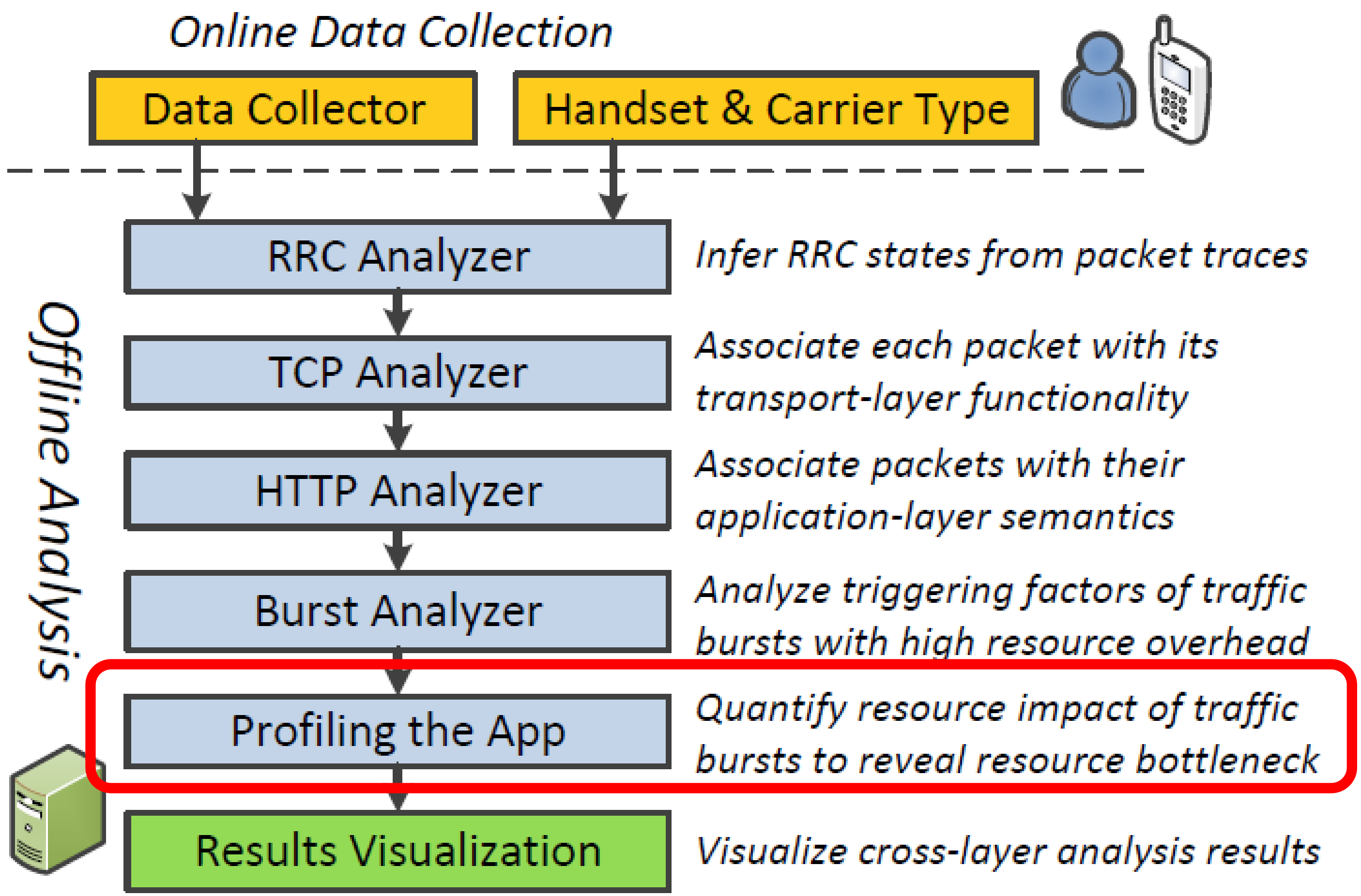


Burst Analysis

- A burst consists of consecutive packets transferred in a batch (i.e., their IAT is less than a threshold)
- We are interested in **short bursts** that incur energy / radio resource inefficiencies
- ARO finds the **triggering factor** of each short burst
  - Triggered by user interaction?
  - By server / network delay?
  - By application delay?
  - By TCP protocol?



ARO System Architecture



Profiling Applications

- From **RRC Analysis**
  - We know the **radio resource state** and the **radio power** at any given time
- From **Burst analysis**
  - We know the **triggering factor** of each burst
  - We know the **transport-layer** and **application-layer** behavior of each burst
- By "**profiling applications**", we mean
  - Compute resource consumption of each **burst**
  - Therefore identify the root cause of resource inefficiency.

Metrics for Quantifying Resource Utilization Efficiency

- Handset radio energy consumption
- DCH occupation time
  - Quantifies **radio resource utilization**
- Total state promotion time (IDLE→DCH, FACH→DCH)
  - Quantifies **signaling overhead**
- Details of computing the three metrics (upperbound and lowerbound) in the paper

Case Studies

- Fully implemented for Android platform (7K LoC)
- Study 17 popular Android applications
  - All in the "TOP Free" Section of Android Market
  - Each has 250,000+ downloads as of Dec 2010
- ARO pinpoints resource inefficiency for many popular applications. For example,
  - **Pandora Streaming**
    High radio energy overhead (50%) of periodic measurements
  - **Fox News**
    High radio energy overhead (15%) due to users' scrolling
  - **Google Search**
    High radio energy overhead (78%) due to real-time query suggestions

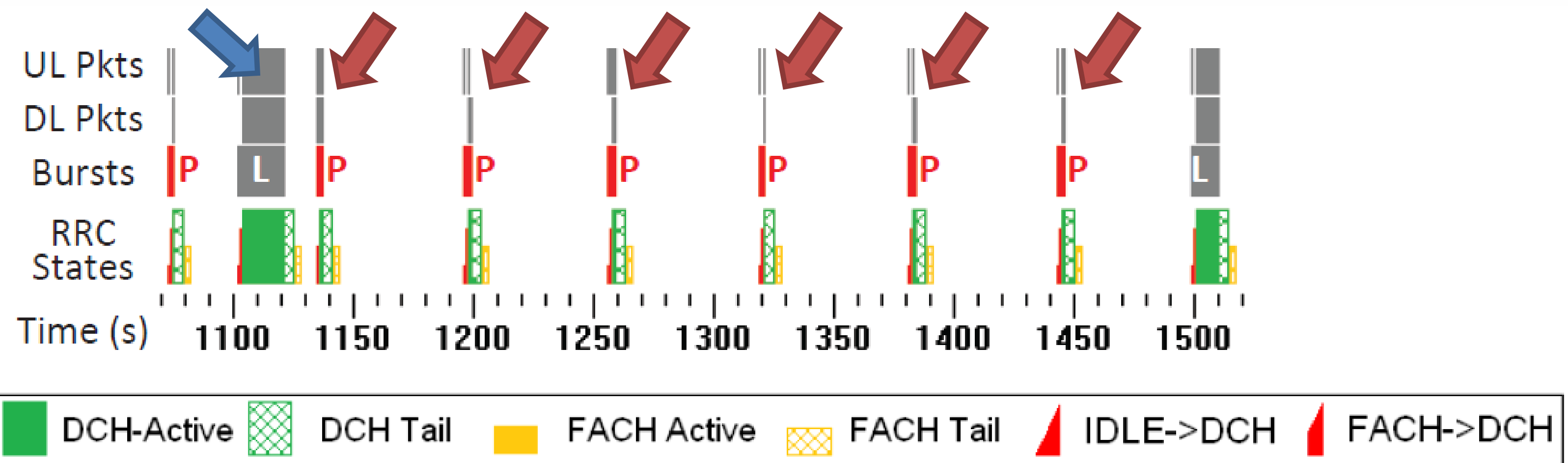
Case Study: Pandora Music

Pandora profiling results (Trace len: 1.45 hours)

Burst type	Payloads	Energy		DCH	
		LB	UB	LB	UB
LARGE_BURST	96.4%	35.6%	35.9%	42.4%	42.5%
APP_PERIOD	0.2%	45.9%	46.7%	40.4%	40.9%
APP	3.2%	12.8%	13.4%	12.4%	12.8%
TCP_CONTROL	0.0%	1.2%	1.6%	1.1%	1.5%
TCP_LOSS_RECOVER	0.2%	0.2%	0.6%	0.3%	0.7%
NON_TARGET	0.0%	1.8%	1.8%	1.7%	1.7%
Total	23.6 MB	846 J		895 sec	

Problem: High resource overhead of periodic audience measurements (every 1 min)

Recommendation: Delay transfers and batch them with delay-sensitive transfers

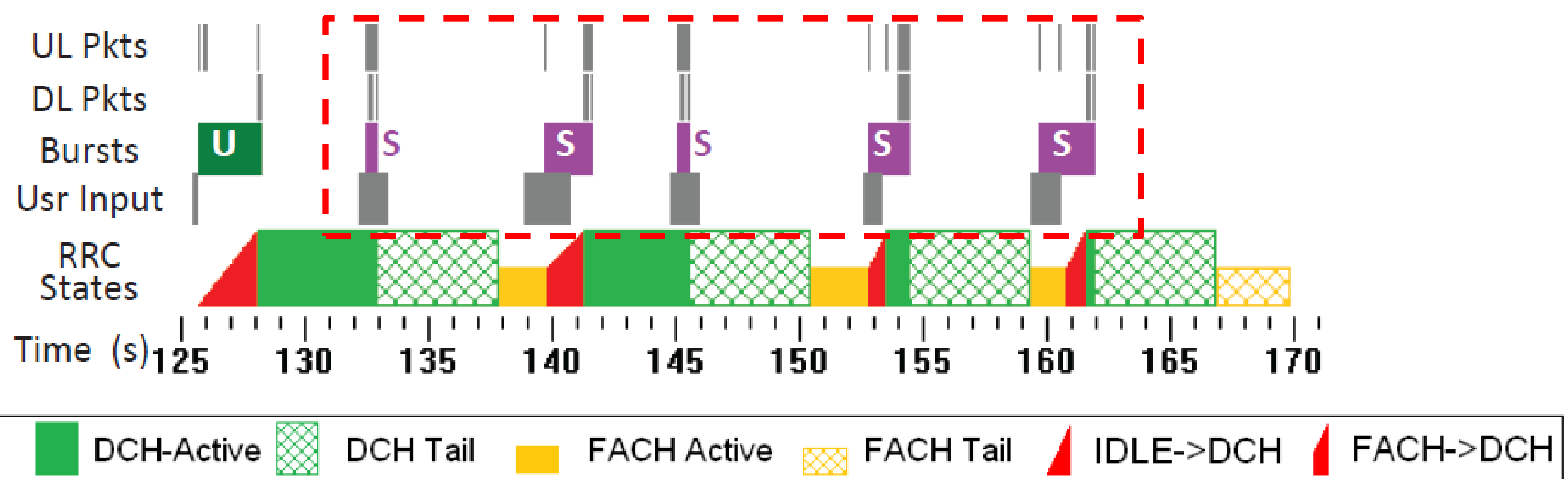


Case Study: Fox News

Fox News profiling results (Trace len: 10 mins)

Burst type	Payloads	Energy		DCH	
		LB	UB	LB	UB
USER_INPUT(Click)	91.0%	56.7%	67.6%	60.2%	70.4%
USER_INPUT(Scroll)	5.9%	15.2%	17.9%	14.7%	16.7%
APP_PERIOD	1.5%	5.2%	7.5%	6.1%	7.4%
TCP_CONTROL	0	0.7%	3.7%	0.0%	2.3%
TCP_LOSS_RECOVER	1.5%	0.7%	2.5%	1.9%	3.2%
SVR_NET_DELAY	0.1%	0.4%	0.8%	0.0%	0.0%
Total	1.0 MB	276 J		284 sec	

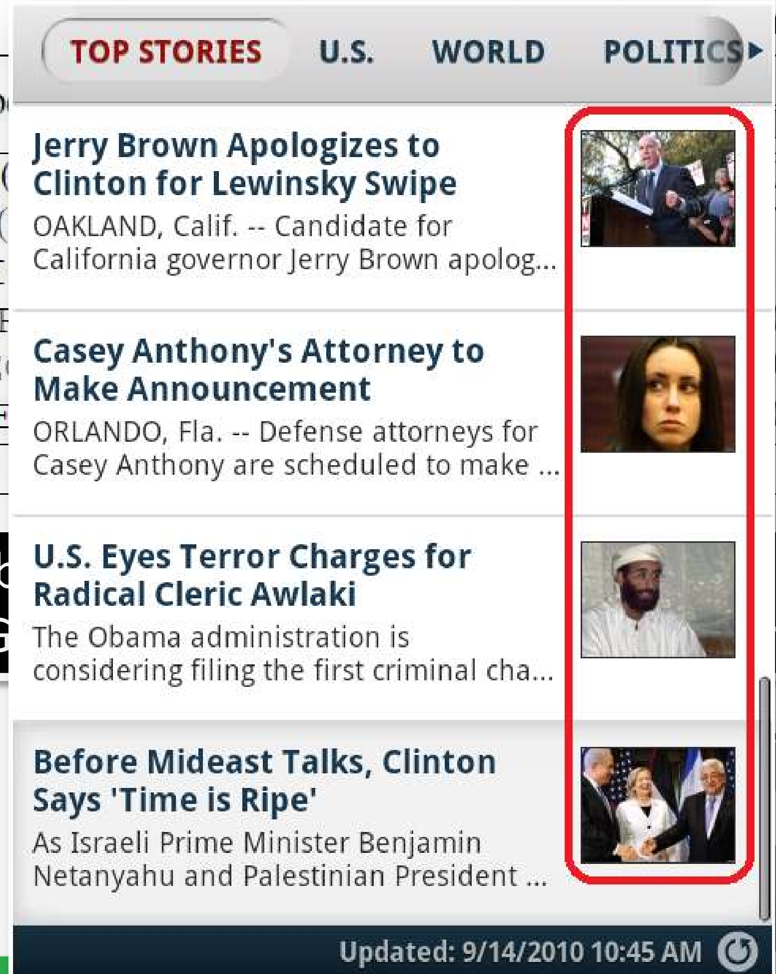
Problem: Scattered bursts due to scrolling
Recommendation: Group transfers of small thumbnail images in one burst



Case Study: Fox News

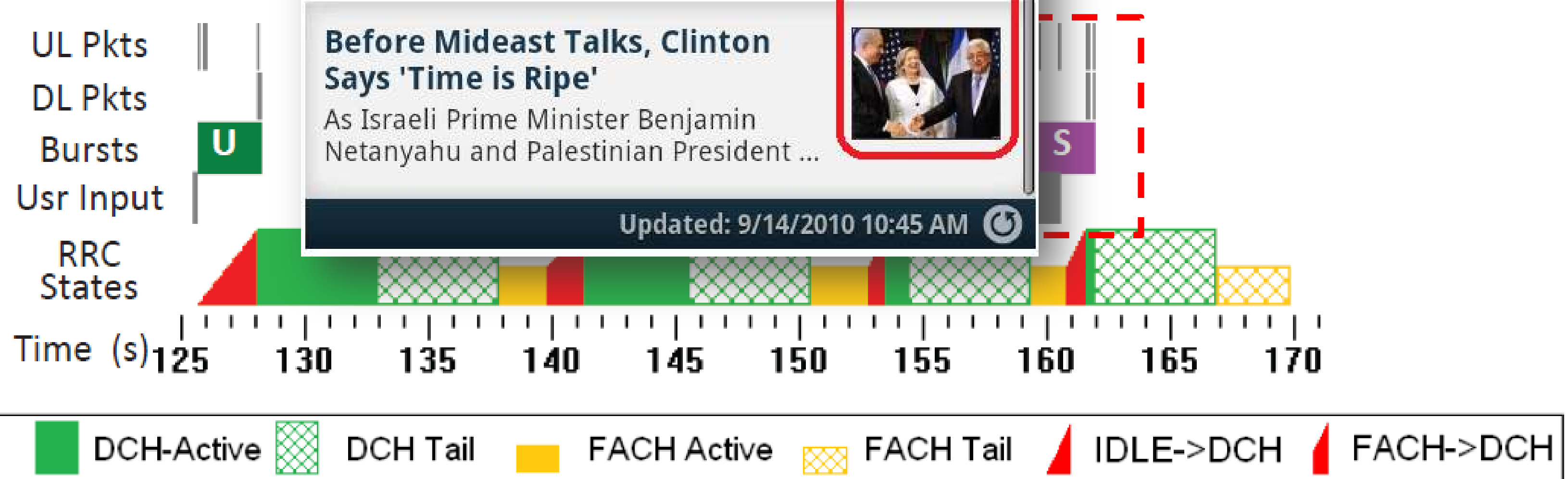
...mins)

Burst typ...	...	DCH	
		...B	UB
USER_INPUT(		...2%	70.4%
USER_INPUT(		...7%	16.7%
APP_PERI		...%	7.4%
TCP_CONTF		...)%	2.3%
TCP_LOSS_RE(		...)%	3.2%
SVR_NET_DE		...)%	0.0%
Total		284 sec	



Problem: Scattered b...

Recommendation: G... ...mages in one burst



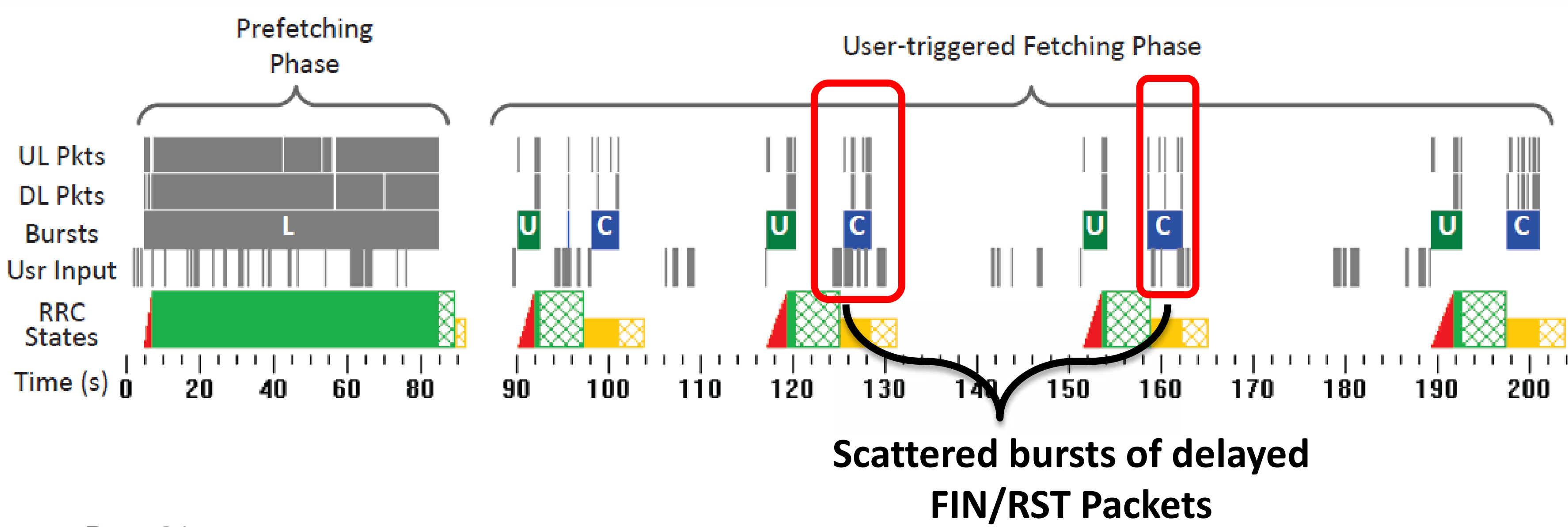
Case Study: BBC News

BBC News profiling results

User-triggered Fetching Phase (8 mins)					
Burst type	Payloads	Energy		DCH	
		LB	UB	LB	UB
TCP_CONTROL	0	11.3%	24.2%	0.0%	5.7%
USER_INPUT	98.7%	42.5%	73.1%	37.9%	90.0%
SVR_NET_DELAY	1%	0.0%	2.7%	0.0%	5.2%
Total	162 KB	145 J		120 sec	

Problem: Scattered bursts of delayed FIN/RST packets

Recommendation: Close a connection immediately if possible, or within tail time



Summary

- ARO helps developers design **cellular-friendly** smartphone applications by providing **visibility** of radio resource and energy utilization.

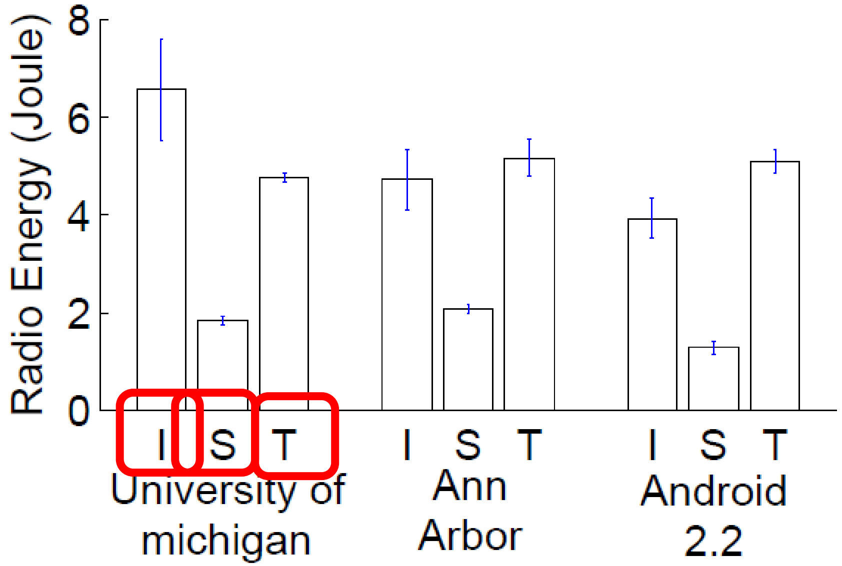
Research Impact

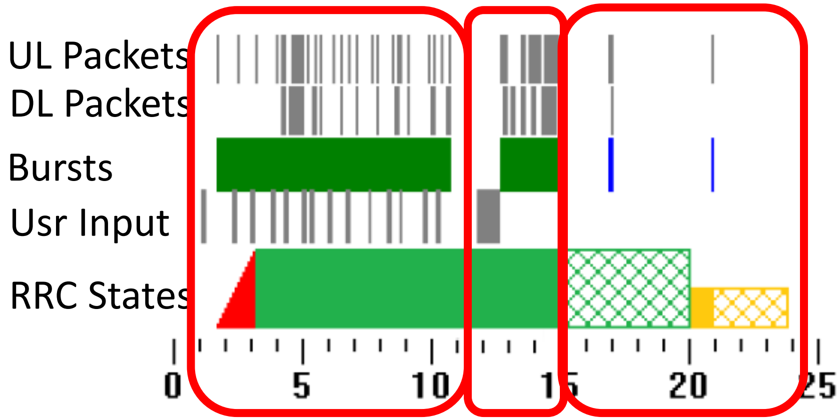
- We contacted developers of top apps (*e.g.*, Pandora). The feedback has been encouragingly positive.

> *AT&T's analysis of the Pandora application gave us a much better view of how Pandora interacts with low-level cellular network resources. Now that we better understand these interactions, we can optimize our application to make more efficient use of these resources. In fact, we'd like to incorporate AT&T's profiling tool as part of our normal ongoing testing.*
>
> *Tom Conrad, CTO of PANDORA®*

- AT&T coverage: "*A Call for More Energy-Efficient Apps*": http://www.research.att.com/articles/featured_stories/2011_03/201102_Energy_efficient
- Production version of ARO is being created.

Case Study: Google Search





Search three key words.

ARO computes energy consumption for three phases

I: Input phase S: Search phase T: Tail Phase

Problem: High resource overhead of query suggestions and instant search

Recommendation: Balance between functionality and resource when battery is low

Case Study: Audio Streaming

Constant bitrate vs. bursty streaming

Name	Server	bitrate	Radio Power
NPR News	SHOUTcast	32 kbps	36 J/min
Tune-in	Icecast	119 kbps	36 J/min
Iheartradio	QTSS	32 kbps	36 J/min
Pandora w/ Ad	Apache	bursty	11.2 J/min
Pandora w/o Ad*	Apache	bursty	4.8 J/min
Slacker	Apache	bursty	10.9 J/min

* A hypothetical case where all periodic ads are removed.

Problem: Low DCH utilization due to constant-bitrate streaming

Recommendation: Buffer data and periodically stream data in one burst