University of Minnesota
Driven to Discover℠

Proactive Wireless Caching for Small Cells in 5G

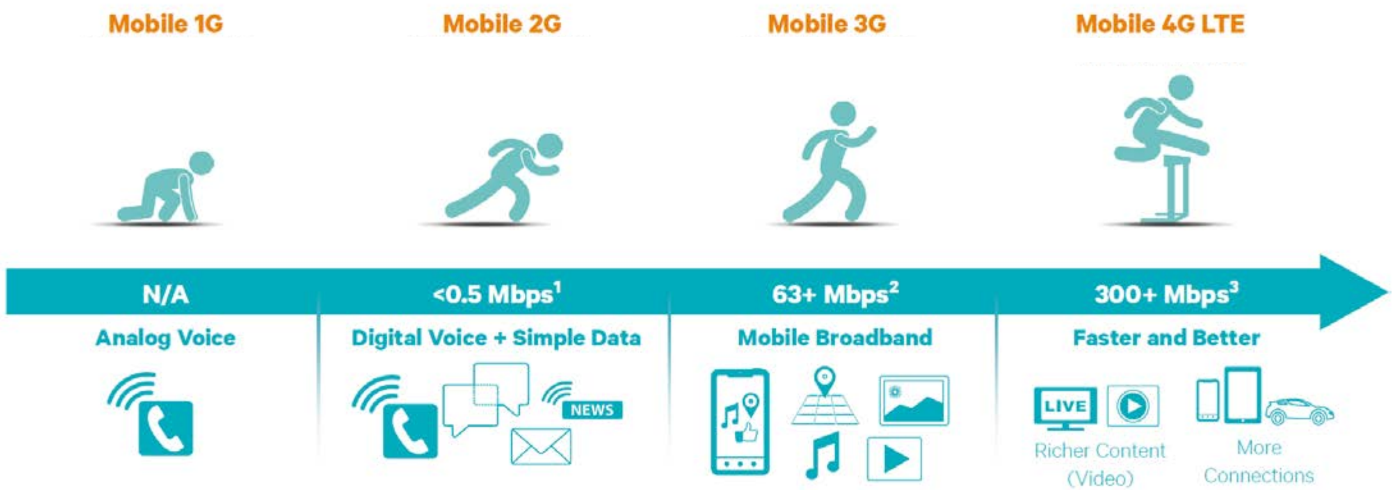
Georgios B. Giannakis

Acknowledgements: *A. Sadeghi and F. Sheikholeslami*
NSF grants 1343248, 1423316, 1508993

PICASSO
EU-US ICT collaboration

Trans-Atlantic Symposium on Technology and Policy for a Smart Society
Minneapolis, MN, June 19, 2017

Evolution of cellular networks



5G

?

?

~1946 ~1980 ~1990 ~2000 ~2009 ~2020

USA
Mobile telephone
(Bell labs)

Japan
Oper. freq. 800 MHz

Finland
TDMA, CDMA
SMS, MMS, picture
High penetration of cell phones
Goodbye 2G by 2018
/1900
Oper. freq. 850/900/1800 MHz
BW: 25 MHz

Japan
3G prompted by basic research
MIMO, STBC, HARQ, OFDM
Mobile Internet via smart phones
Video conference, video on demand
Location based services – GPS
Oper. freq. 800/850/900/1700/1900/2100 MHz
BW: 60 MHz

Norway
HD mobile TV, gaming, amended mobile web access
OFDMA, SC-FDMA, channel dependent scheduling, link adaptation, mobile IP, MU-MIMO, spatial multiplexing
Oper. freq. 700/800/850/900/1700/1800/1900/2100/2300/2500/2600 MHz
BW: 100MHz

- New generation every 10 years: Higher rates, but not backward compatible

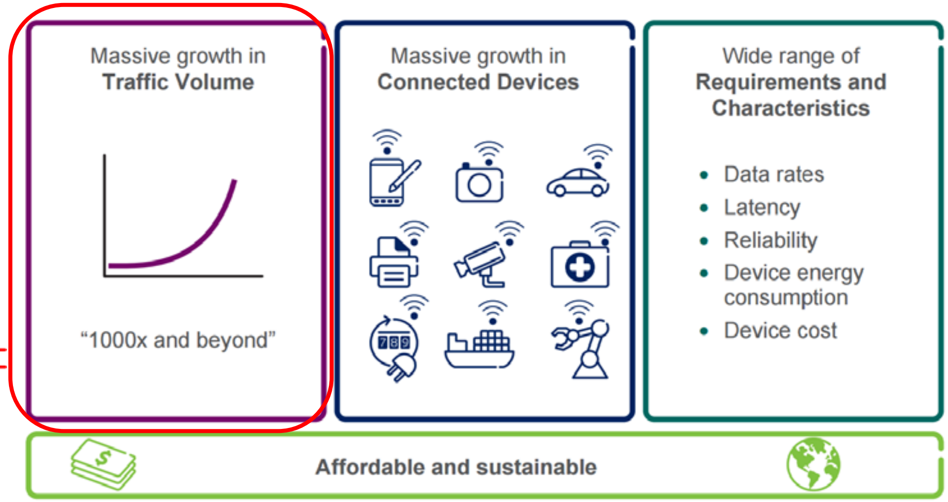
Technology directions for 5G

1. Device-centric designs (edge)
2. Millimeter wave (mmWave)
3. Massive MIMO
4. Smart network, e.g., adaptive caching in small cells (fog)
5. Seamless machine-to-machine (M2M), and IoT communications

Subject of today's talk

F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communication Magazine*, vol. 52, no. 2, pp. 74-80, February 2014.

Challenges for 5G

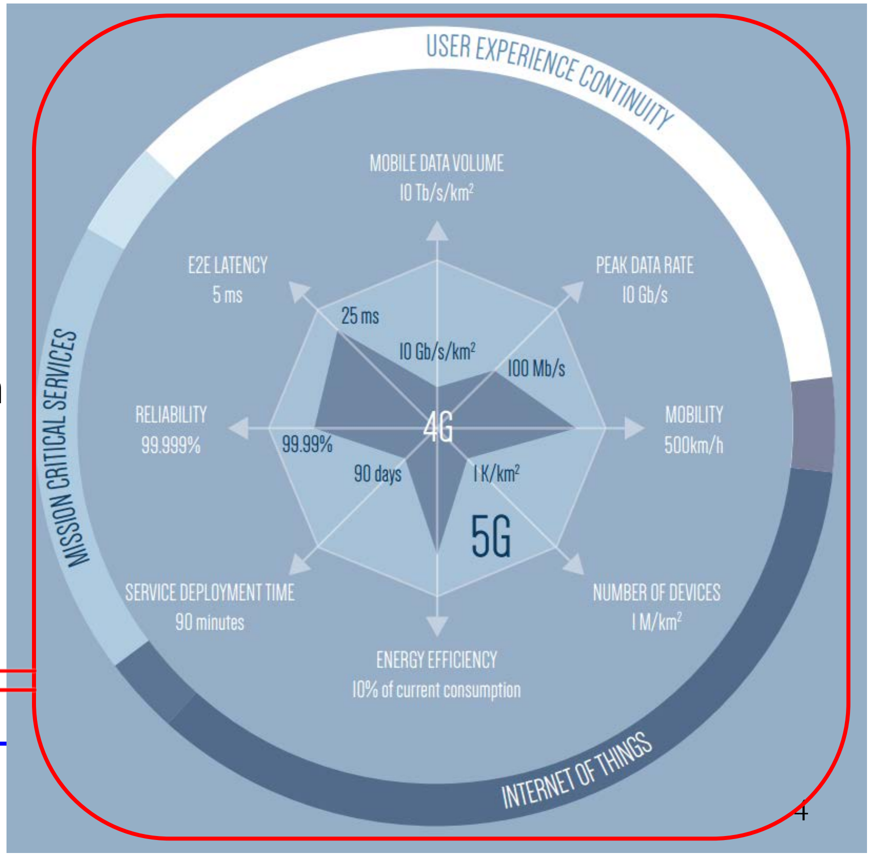


- Evolving user content profiles
  - Mobile video streaming
    - ✓ 50% of traffic volume in 4G
    - ✓ 500-fold increase in 5G
  - Social networking
    - ✓ 15% of traffic volume in 4G
  - Music, games …

5G desiderata over 4G

- Higher volume, lower energy consumption
- Massive number of devices (M2M, IoT)
- Service deployment time

Heterogeneous (small cell) networks (fog)



Heterogeneous small-cell networks

❑ Traditional cells

- Medium to long-range (1-10 km)
- High-gain antenna
- Crucial for coverage and mobility support
- Expensive (over $100K+OpEx)
- 40W Tx power
- Fast dedicated backhaul

❑ Pico-cells

- Short-range (~100m)
- Small, easy deployment
- Targeting “hotspots” or dense areas
- Low-cost ($5-40K + small OpEx)
- 1-2W Tx power
- Cheap backhaul

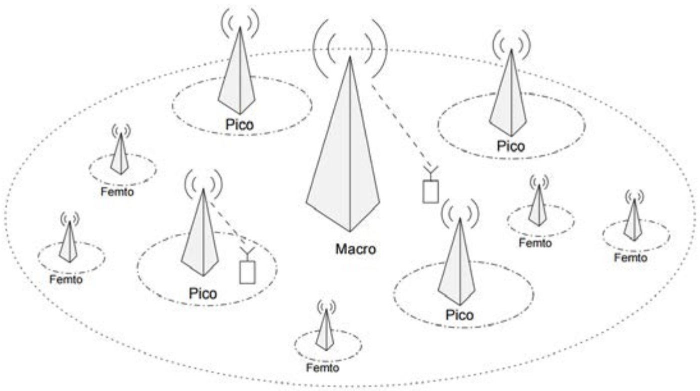
❑ Femto-cells

- WiFi range (~10m)
- Inter-cell coordination minimum overhead
- Licensed spectrum
- 100$+small OpEX
- 100-200mW Tx power
- Backhaul (IP, e.g. DSL, coax)

Heterogeneous networks (HetNets) entail all three types

J. G. Andrews, H. Claussen, M. Dohler, S. Rangan and M. C. Reed, “Femtocells: Past, Present, and Future,” *IEEE Jour. on Selected Areas in Communications*, vol. 30, no. 3, pp. 497-508, April 2012.

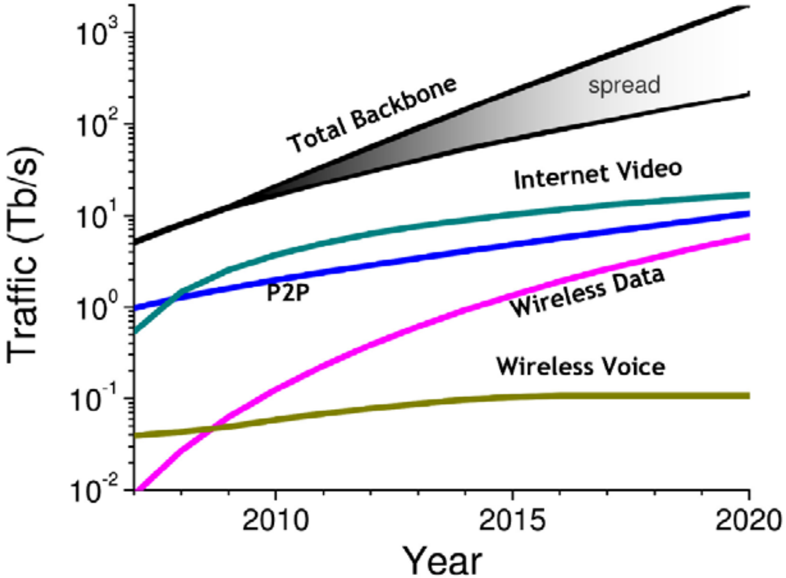
Pros and cons of HetNets



\+ Spatial reuse boosts rate (1000x)

\+ Densification enhances coverage

\+ Reduced cost and energy efficient



\- Backhauling

- Fiber to pico-cells not viable
- Cheap backhauling lowers traffic

\- Increased interference

\- Increased overhead (small cell coordination)

Reusable content overloads backhaul (now 60% of mobile data traffic, 500X in 10yrs)

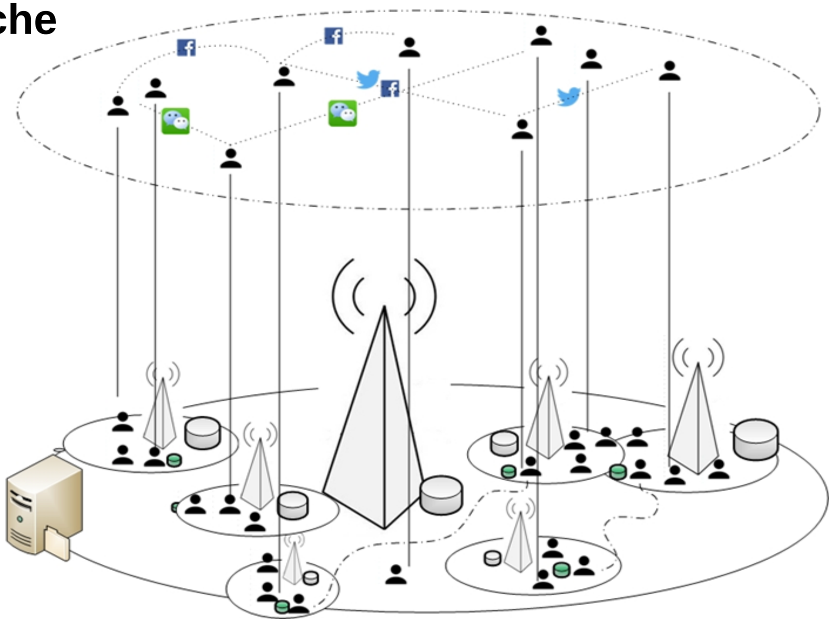
V. Chandrasekhar, J. G. Andrews, and A. Gatherer, “Femtocell networks: A survey,” *IEEE Communication Magazine*, vol. 46, no. 9, pp. 59–67, September 2008.

Proactive caching for small cells

- **Small cells with storage and cashing**
  - Store (coded) popular reusable content; also dynamically via mobile devices
  - Reduce peak-to-average load ratio; and shift backhaul load via caching

- **Pre-allocation of resources**
  - Serve predictable peak-hour demand during off-peak times

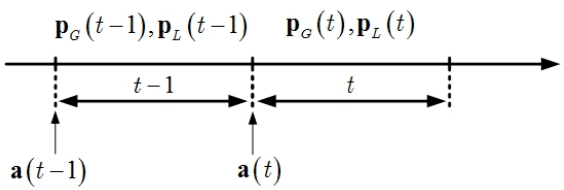
- **Learn and track what, when, and where to cache**
  - Model and learn content popularities
    - ✓ vector of probabilities $[\mathbf{p}]_f = 1/(\alpha f^{\beta})$
    - ✓ popularity matrix – both static models
    - ✓ learn via multi-arm bandit [Belasco etal'14]
  - Account for space-time dynamic changes?

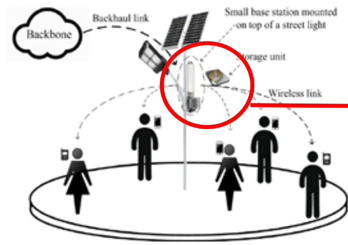


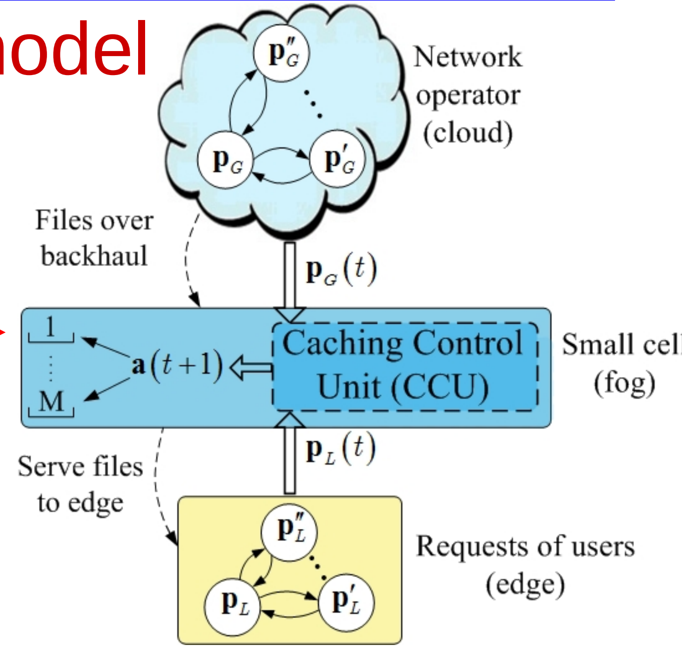
- Leverage inter-dependent social networks?

G. Paschos, E. Bastug, I. Land, G. Caire and M. Debbah, "Wireless caching: Technical misconceptions and business barriers," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 16-22, August 2016.

Space-time dynamic popularity model







- Storage unit cashing *M* (out of total *F*) files
- Local and global profiles per slot *t*

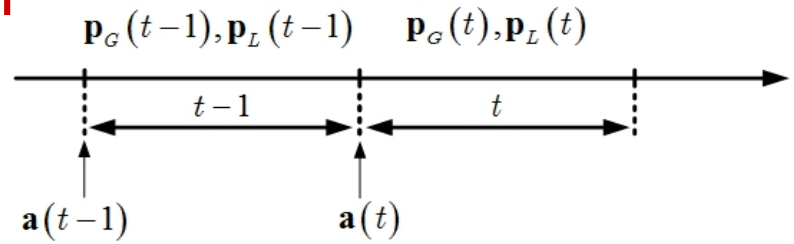
$$[\mathbf{p}_\mathrm{L}(t)]_f = \frac{\text{local requests for } f}{\text{all local requests}} \qquad [\mathbf{p}_\mathrm{G}(t)]_f = \frac{\text{global requests for } f}{\text{all global requests}}$$

- Action vector *F*x1: $[\mathbf{a}(t)]_f = 1$, if file *f* is cashed; 0 otherwise
- State vector 3*F*x1: $\mathbf{s}^\top(t) := [\mathbf{p}_L(t), \mathbf{p}_G(t), \mathbf{a}(t)]$
- Policy $\pi(\cdot)$: $\mathbf{a}(t+1) = \pi(\mathbf{s}(t))$

Goal: Given $\{\mathbf{s}(\tau)\}_{\tau=0}^{t}$ and observed costs, optimize policy interactively

A. Sadeghi, F. Sheikholeslami, and G. B. Giannakis, "Optimal Dynamic Proactive Caching via Reinforcement Learning," *Proc. of Globecom Conference*, Singapore, Dec. 4-8, 2017.

Reinforcement learning approach



- Cost is a weighted combination of three factors
  - Cost of extra files to cache at slot *t*
  - Cost of locally requested non-cached files
  - Cost of globally popular non-cashed files

$$c_1\left(\mathbf{a}(t), \mathbf{a}(t-1)\right) := \lambda_1 \mathbf{a}^\top(t)(\mathbf{1} - \mathbf{a}(t-1))$$

$$c_2\left(\mathbf{s}(t)\right) := \lambda_2 (\mathbf{1} - \mathbf{a}(t))^\top \mathbf{p}_L(t)$$

$$c_3\left(\mathbf{s}(t)\right) := \lambda_3 (\mathbf{1} - \mathbf{a}(t))^\top \mathbf{p}_G(t)$$

$$C\left(\mathbf{s}(t-1), \mathbf{a}(t) | \, \mathbf{p}_{\mathrm{G}}(t), \mathbf{p}_{\mathrm{L}}(t)\right) := c_1\left(\mathbf{a}(t), \mathbf{a}(t-1)\right) + c_2\left(\mathbf{s}(t)\right) + c_3(\mathbf{s}(t))$$

- Expected cost (discounted by 0<*γ*<1) value function

$$V^\pi\left(\mathbf{s}(t_0)\right) := \lim_{T\to\infty} \mathbb{E}\left[\sum_{t=t_0+1}^{T} \gamma^{t-t_0-1} C\left(\mathbf{s}(t-1), \pi\left(\mathbf{s}(t-1)\right) | \mathbf{p}_{\mathrm{G}}(t), \mathbf{p}_{\mathrm{L}}(t)\right)\right]$$

$$\pi^*(\mathbf{s}_0) = \arg\min_{\pi\in\Pi} V^\pi\left(\mathbf{s}_0\right), \ \forall \mathbf{s}_0$$

- Possible solvers
  - ✓ Adaptive dynamic programming ✓ SARSA ✓ Q-learning

R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, Cambridge, MA, USA: MIT Press, 1998.

Q-learning

- Markov decision process model $\mathbf{s}^\top(t) := [\mathbf{p}_L(t), \mathbf{p}_G(t), \mathbf{a}(t)]$
  - ✓ Transition per action $[\boldsymbol{P}^a]_{ss'} = \Pr(\mathbf{S}(t+1) = \mathbf{s}' | \mathbf{S}(t) = \mathbf{s}, \mathbf{A}(t) = \mathbf{a})$
  - ✓ (state, action) value function $Q^\pi(\mathbf{s}, \mathbf{a}) = \mathbb{E}\left[C(\mathbf{s}, \mathbf{a} | \mathbf{p}_L, \mathbf{p}_G)\right] + \gamma \sum_{\mathbf{s}'} [\boldsymbol{P}^a]_{\mathbf{ss}'} V^\pi(\mathbf{s}')$
  - Initialize $\mathbf{s}(0),\ \hat{Q}(\mathbf{s}, \mathbf{a}) = 0$

S1. Action (**exploration** or **exploitation**)

$$\mathbf{a}(t) = \begin{cases} \arg\min_{\mathbf{a}} \hat{Q}\left(\mathbf{s}(t-1), \mathbf{a}\right) & \text{w.p.} \quad 1 - \epsilon_t \\ \mathbf{a} \in \mathcal{A} & \text{w.p.} \quad \epsilon_t \end{cases}$$

S2. Update $\mathbf{p}_{\mathrm{L}}(t),\ \mathbf{p}_{\mathrm{G}}(t)$ based on user requests

S3. Cost $C\left(\mathbf{s}(t-1), \mathbf{a}(t) | \, \mathbf{p}_{\mathrm{G}}(t), \mathbf{p}_{\mathrm{L}}(t)\right)$

S4. Q-table update

$$\hat{Q}_t\left(\mathbf{s}(t-1), \mathbf{a}(t)\right) = (1 - \beta_t)\hat{Q}_{t-1}\left(\mathbf{s}(t-1), \mathbf{a}(t)\right) + \beta_t \times \left[C\left(\mathbf{s}(t-1), \mathbf{a}(t) | \, \mathbf{p}_{\mathrm{G}}(t), \mathbf{p}_{\mathrm{L}}(t)\right) + \gamma \min_{\mathbf{a}} \hat{Q}_{t-1}\left(\mathbf{s}(t), \mathbf{a}\right)\right]$$

- Prohibitive complexity motivates Q-function approximation alternatives

Theorem. For any $\epsilon_t > 0$, if $\sum_{t=1}^{\infty} \beta_t = \infty$ and $\sum_{t=1}^{\infty} \beta_t^2 < \infty$, then $\pi \to \pi^*$ w.p.1

Preliminary tests

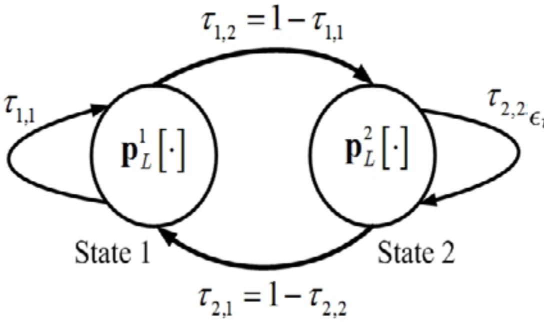
- *F*=10 files and *M*=2 storage units
- Two-state Markov chain for modeling $\mathbf{p}_{\mathrm{L}}[t]$
- State transition probabilities

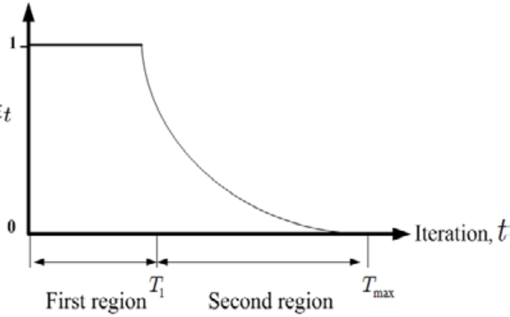
$$\boldsymbol{\tau} := \begin{bmatrix} 0.35 & 0.65 \\ 0.75 & 0.25 \end{bmatrix}$$

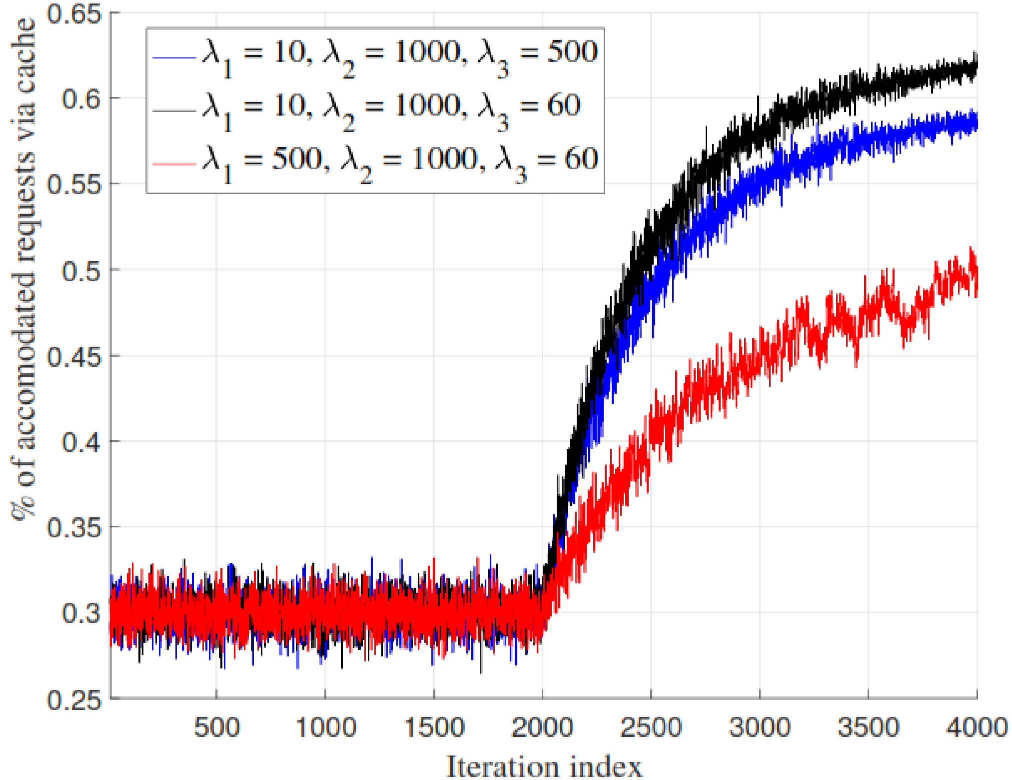
- Similarly for $\mathbf{p}_{\mathrm{G}}[t]$

$$\boldsymbol{\tau}' := \begin{bmatrix} 0.6 & 0.4 \\ 0.45 & 0.55 \end{bmatrix}$$

- A two-step structure for ϵ_t
- $\beta_t = 0.1$ and $\gamma = 0.6$



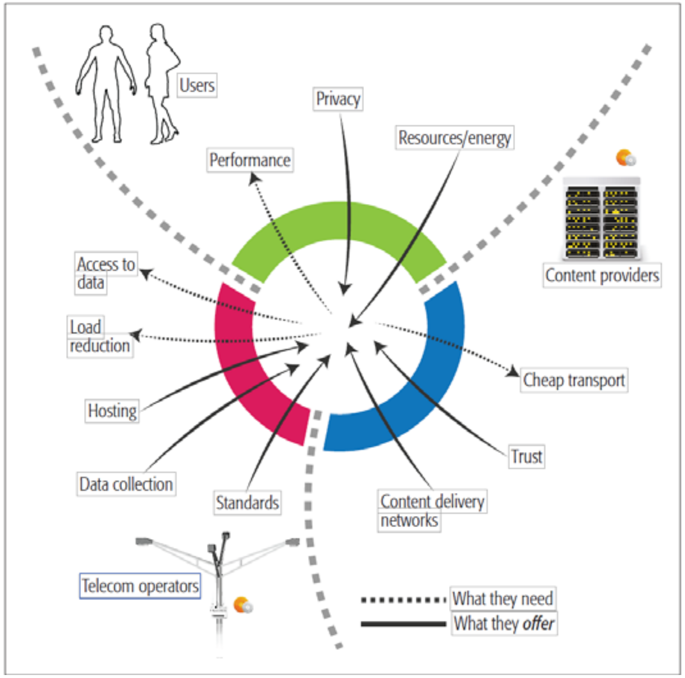




A. Sadeghi, F. Sheikholeslami, and G. B. Giannakis, “Optimal Dynamic Proactive Caching via Reinforcement Learning,” *Proc. of Globecom Conference*, Singapore, Dec. 4-8, 2017.

Future research and stakeholder analysis

- ❑ Low-complexity tracking of dynamic content popularities
- ❑ Cooperative cashing across neighboring small cells
- ❑ Cross-layer design of coded cashing
- ❑ Privacy-preserving, secure, space-time variable caching

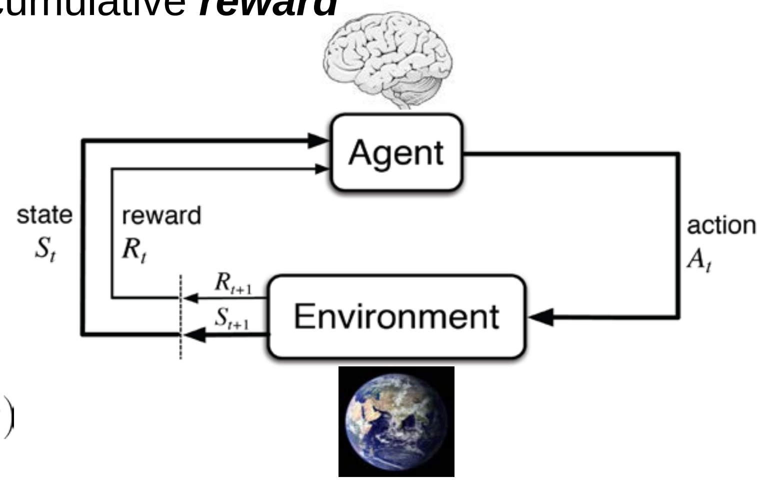


Thank you!

G. Paschos, E. Bastug, I. Land, G. Caire and M. Debbah, “Wireless caching: Technical misconceptions and business barriers,” *IEEE Communications Magazine*, vol. 54, no. 8, pp. 16-22, August 2016.

Reinforcement learning

- Agent learns how to optimize objective by interacting with an unknown environment
- **Objective:** Find optimal policy maximizing expected cumulative ***reward***



- Agent-environment interactions entail
  - States (***S***); actions (***A***); and rewards (***R***)

- **Model:** Markov decision process (MDP) $(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{R}, \boldsymbol{P}, \gamma)$
  - Transition matrix per action *a*: $[\boldsymbol{P}^a]_{ss'} = \Pr(S_{t+1} = s' | S_t = s, A_t = a)$

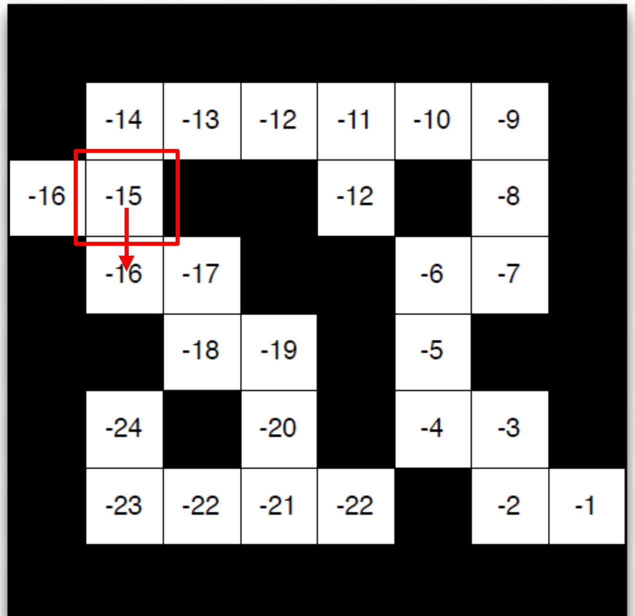
- MDP value functions for policy $\pi(\cdot)$, which maps state to action $\mathbf{a}[t+1] = \pi(\mathbf{s}[t])$
  - (State) value function $V^\pi(s) = \mathbb{E}[\sum_{\tau=0}^{\infty} \gamma^\tau R^\pi_{t+\tau+1} | S_t = s]$ $\gamma \in [0, 1]$
  - (State,action) value function $Q^\pi(s, a) = \mathbb{E}[R^a_{t+1} + \sum_{\tau=1}^{\infty} \gamma^\tau R^\pi_{t+\tau+1} | S_t = s, A_t = a]$

R. S. Sutton, and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 2014.

RL example: Maze runner robot

Objective: Exit the maze ASAP

Start



Goal

State: agent's coordinates (square)

Actions: move N, E, S, W

Reward: -1 per step taken to exit

- Here *s*, *a*, and *R* are deterministic

Goal of RL → Arrows: optimal policy, $\pi^*(s), \ \forall s$

→ **Value function**: $V^{\pi^*}(s) = -15$

→ **Q-function:** $Q^{\pi^*}(s, a = \mathrm{S}) = -17$

- Policy π^* is optimal iff for $\forall \pi$:

$$\boxed{V^{\pi^*}(\mathbf{s}) \geq V^{\pi}(\mathbf{s}); \quad \forall \mathbf{s} \in \mathcal{S}}$$

$$\pi^*(\mathbf{s}) = \arg\max_{\mathbf{a}} Q^{\pi^*}(\mathbf{s}, \mathbf{a})$$
$$V^{\pi^*}(\mathbf{s}) = \max_{\mathbf{a}} Q^{\pi^*}(\mathbf{s}, \mathbf{a})$$