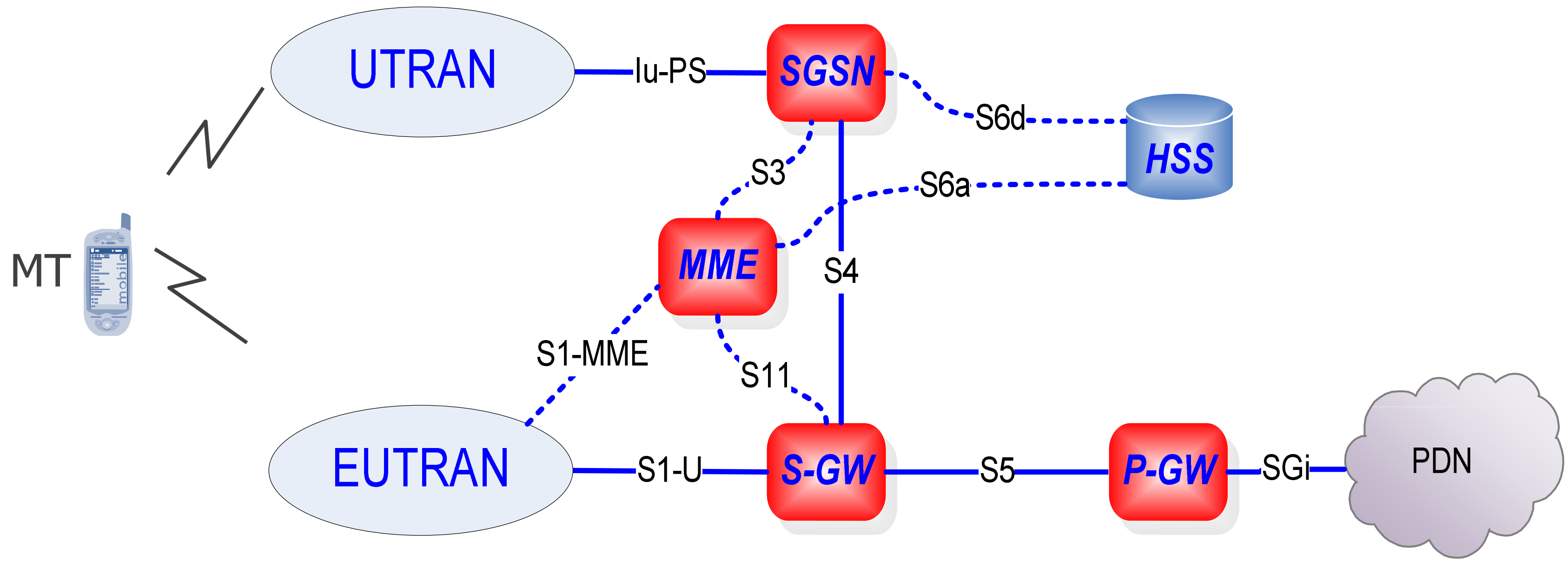
Mitigating Signalling Overhead from Multi-Mode Mobile Terminals

Indra Widjaja and Carl Nuzman

Bell Labs, Alcatel-Lucent

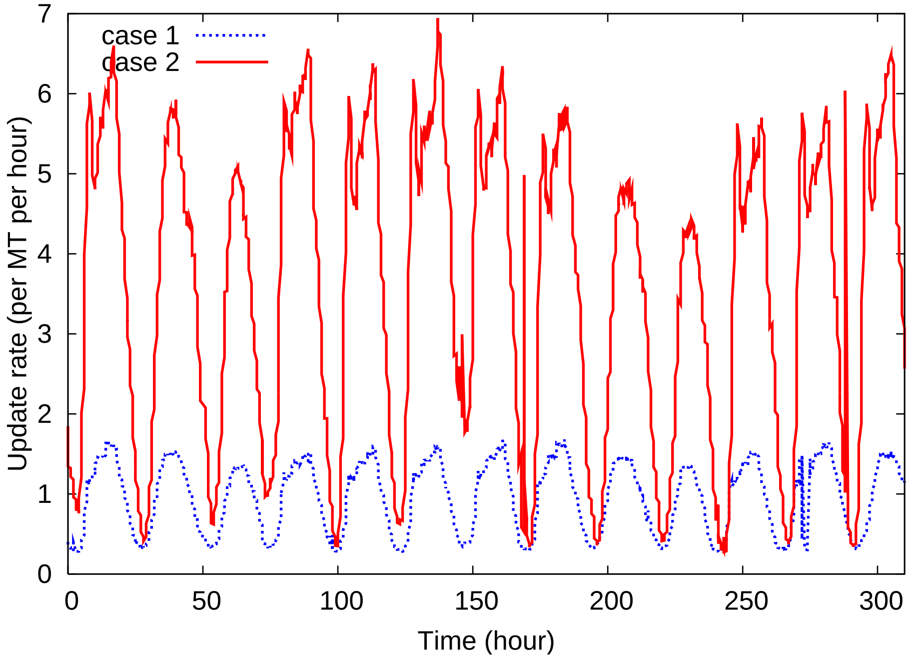
Alcatel·Lucent

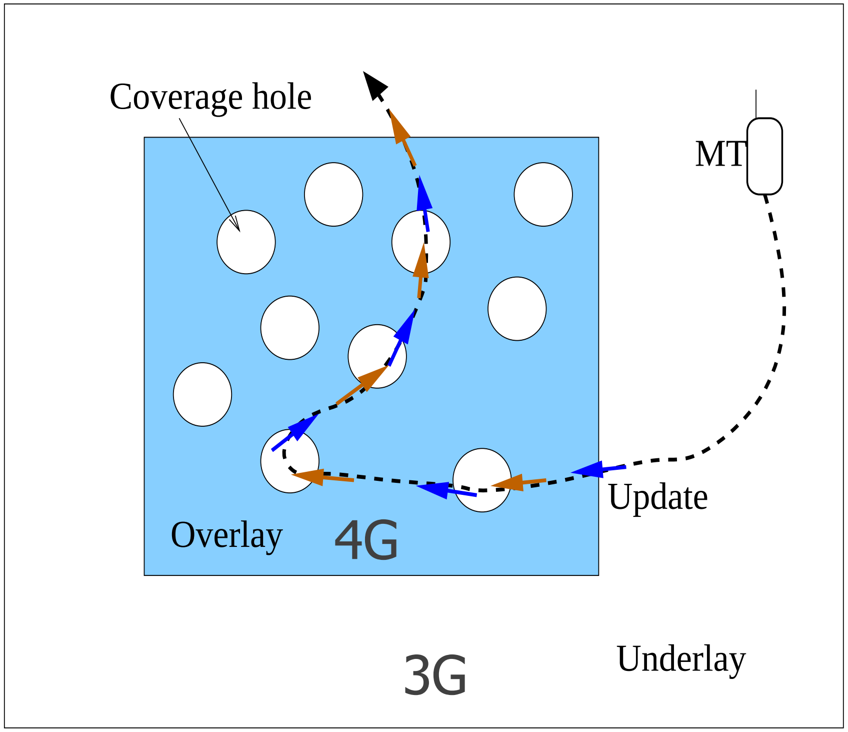
Background



- Multiple technologies (2G/3G/4G) are increasingly deployed simultaneously by cellular-network operators.
- Mobile terminals (MTs) have multi-mode capability and can switch from one technology to another *seamlessly*.

Problem and Motivation

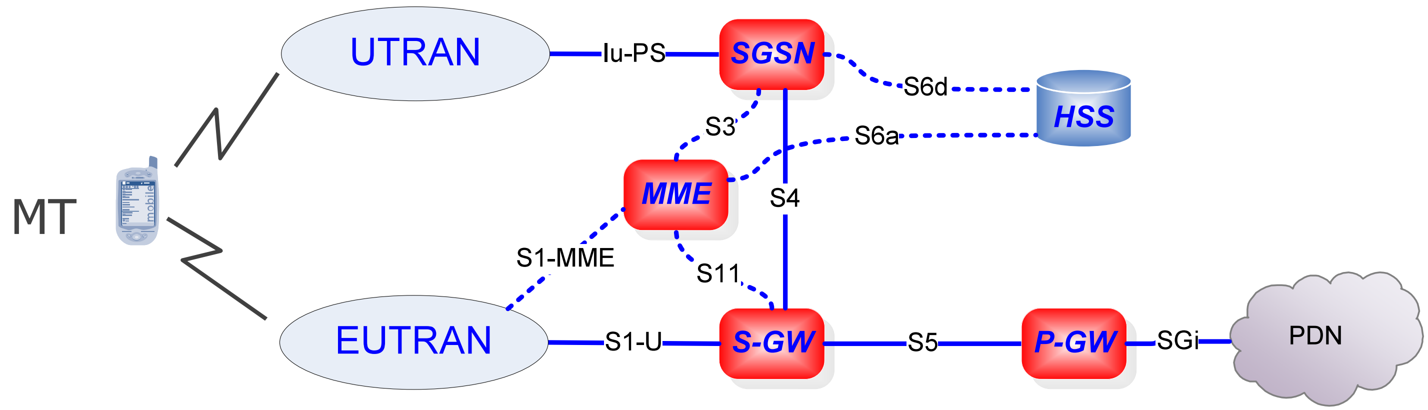




Frequent updates can stress the control plane of the network

- Can we characterize the update rate due to registration ping-ponging?
- Can we minimize the impact on signaling load?

Solution from Standards - Idle-state Signaling Reduction (ISR)



When **ISR is not activated**, MT is registered with either technology:

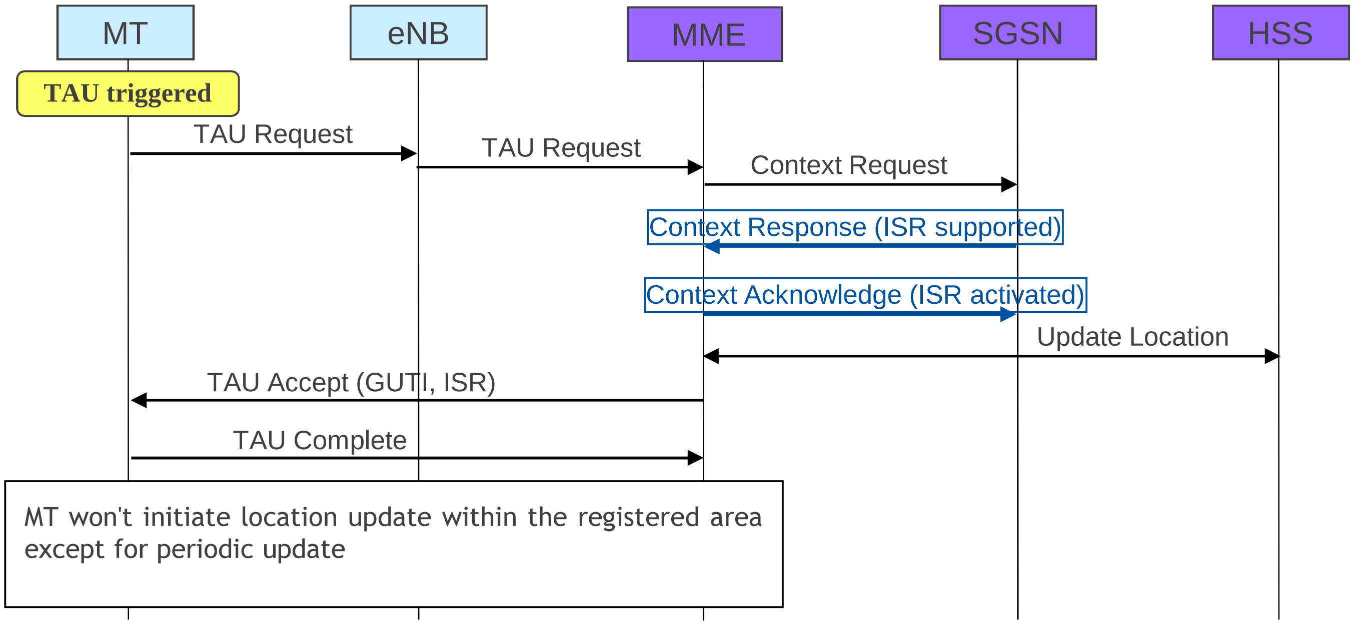
- An MT moving from one cell of a given technology to another cell of a different technology has to perform a location update (TAU/RAU).

When **ISR is activated**, MT is registered with both technologies:

- No update is triggered when an MT moves from one cell to another cell of a different technology.

ISR is currently perceived to be a good thing

ISR Activation Example



- During an update, the network decides whether to activate ISR individually for each MT
- Once ISR is activated, it remains active until the network decides not to re-activate during the next update
- MT and the network run periodic update timers. The network performs implicit detach which deactivates ISR if it does not receive a periodic update after the timer expires. MT deactivates ISR if it cannot perform a periodic update.

Analysis of Update Rate

- Mobility model
  - MT location is uniform with density ρ
  - Direction of travel is uniformly distributed over [0, 2π]
  - In a closed region with perimeter length *L*, the average rate MTs with velocity *V* cross the perimeter is

$$R = \frac{\rho V L}{\pi}$$

  - The average update rate per MT without ISR can be expressed by

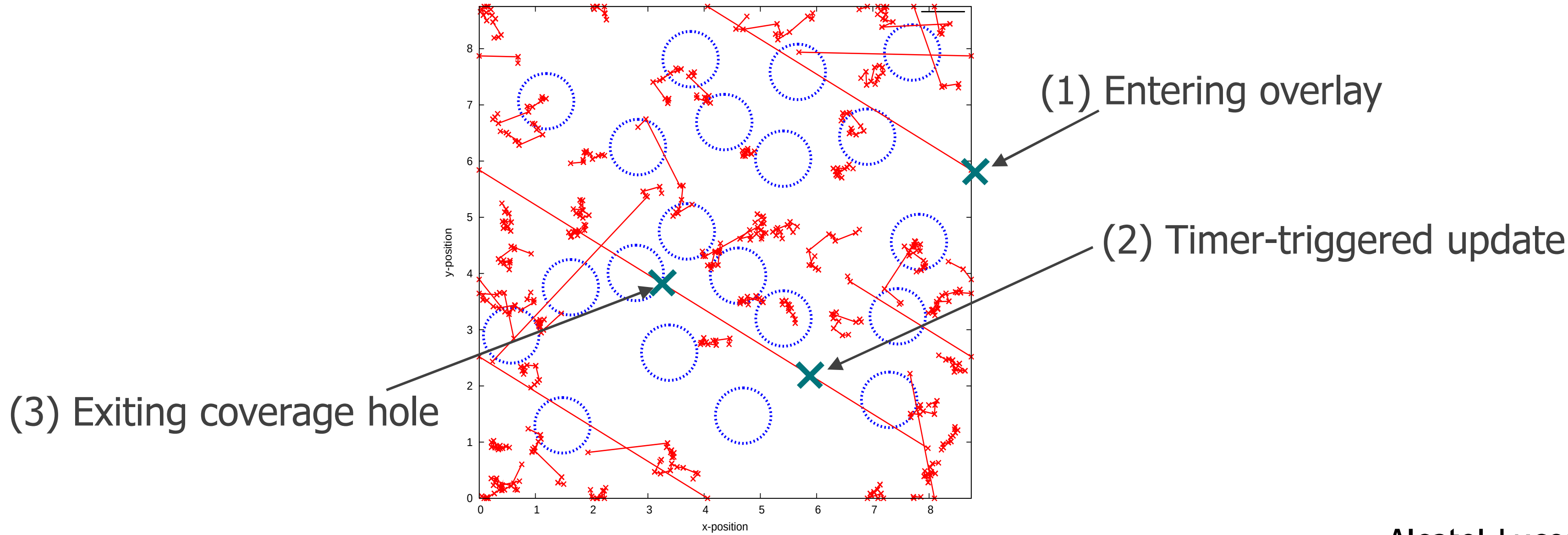
$$\hat{R} = \left(\frac{C_o}{C_a}\right)\frac{\rho V L_a}{\pi A o}$$ (1) due to MT entering an overlay region

$$+\frac{\mu}{e^{\mu T}-1}$$ (2) due to timer-triggered updates (period T) while in overlay

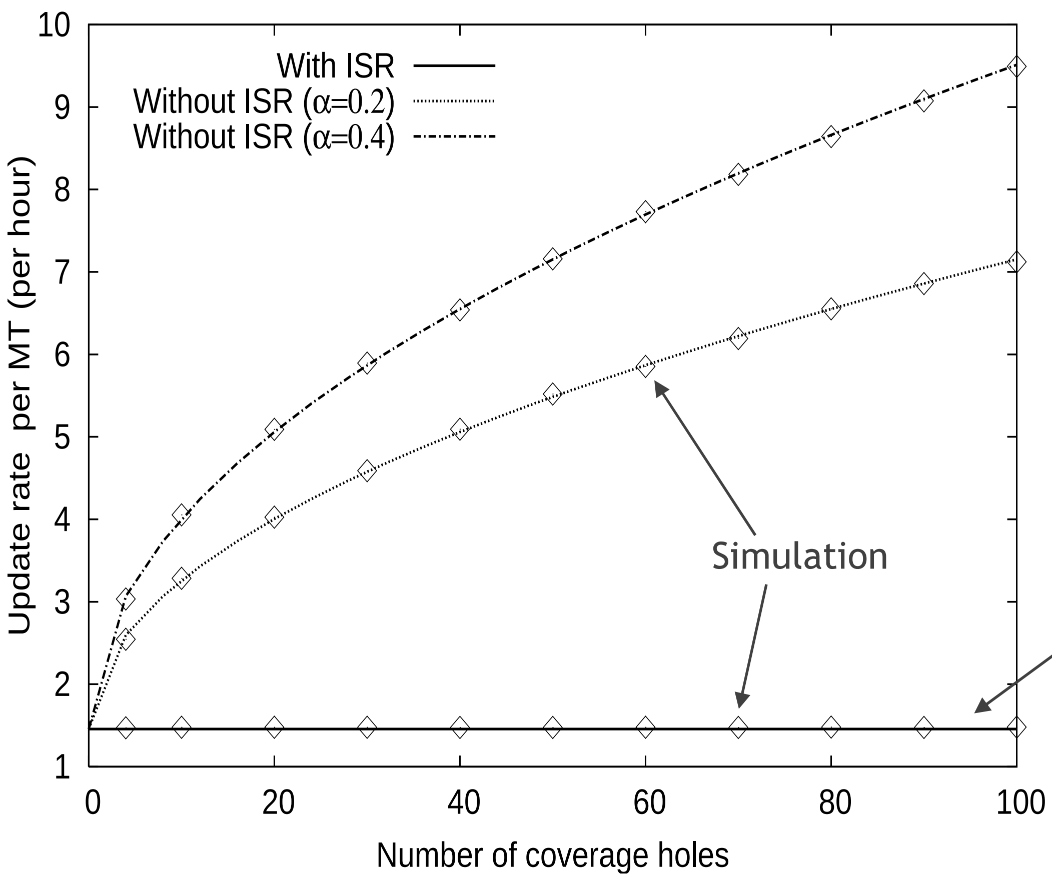
$$+2V\sqrt{\frac{\alpha N_h}{\pi A_o}}$$ (3) due to MT exiting overlay coverage holes

Simulation of Update Rate

- Random waypoint on a torus
  - MT's journey consists of a sequence of flights with a pause between two consecutive flights.
  - In each flight, the duration and velocity are picked independently according to given density functions, and flight direction is uniformly distributed over [0, 2π]. The path of each flight follows a straight line.
  - The pause time is independently chosen according to a given density function.
  - When MT hits the boundary of a rectangular region, it is wrapped at the opposite side.



Average Update Rate per MT



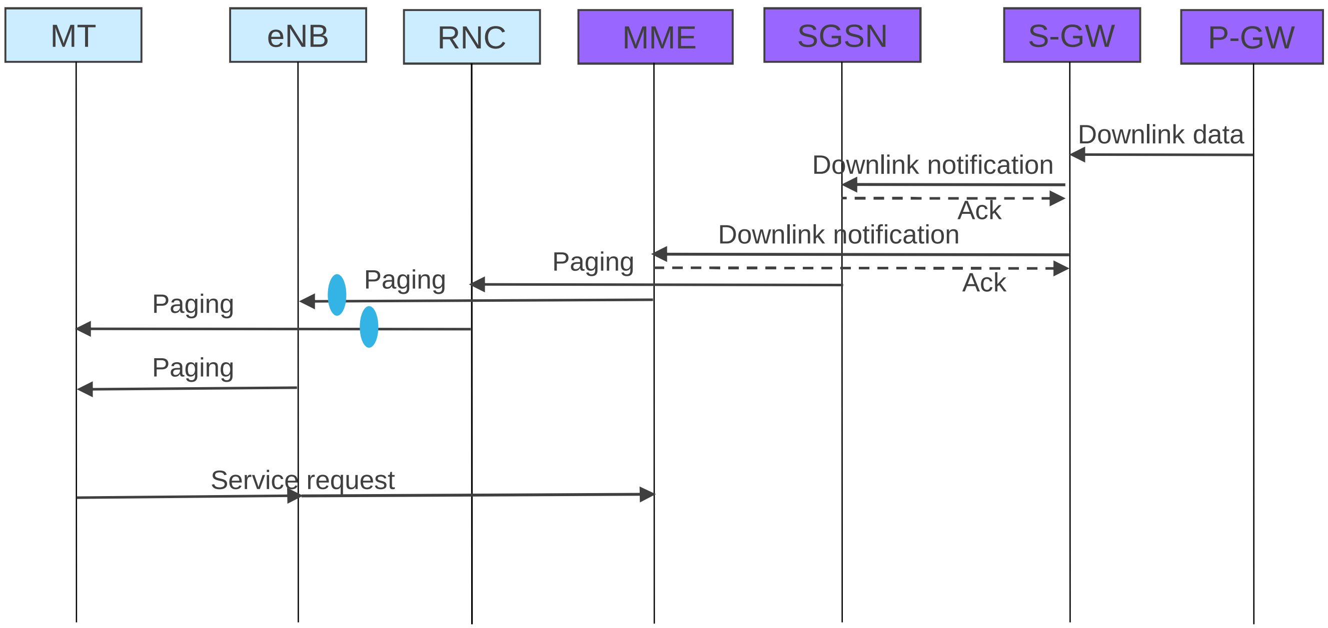
Update rate grows as a square root of number of coverage holes without ISR for a fixed total hole area

ISR is beneficial

Assumptions:

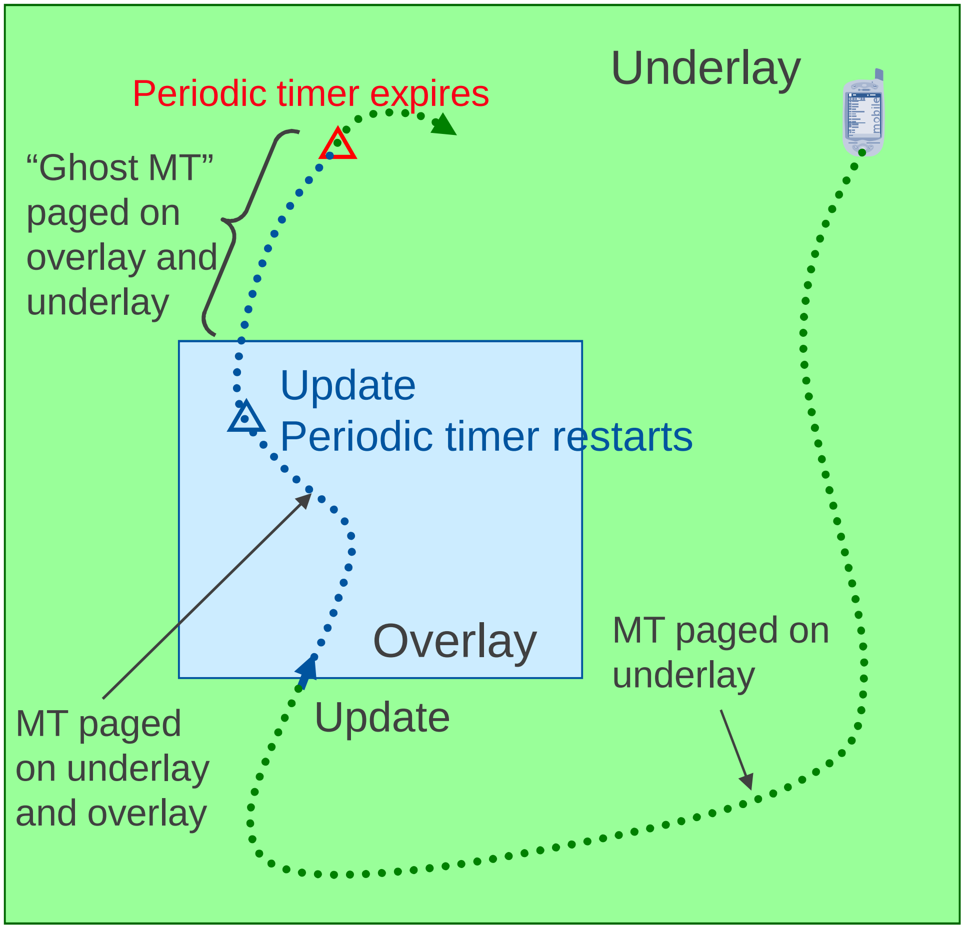
- Overlay size = 100 eNBs, *V*=10 km/hr, Cell perimeter length = 3.5 km, Periodic timer = 3 hrs
- Note: α is the ratio of total area of coverage holes to overlay area

What Happens with Paging?



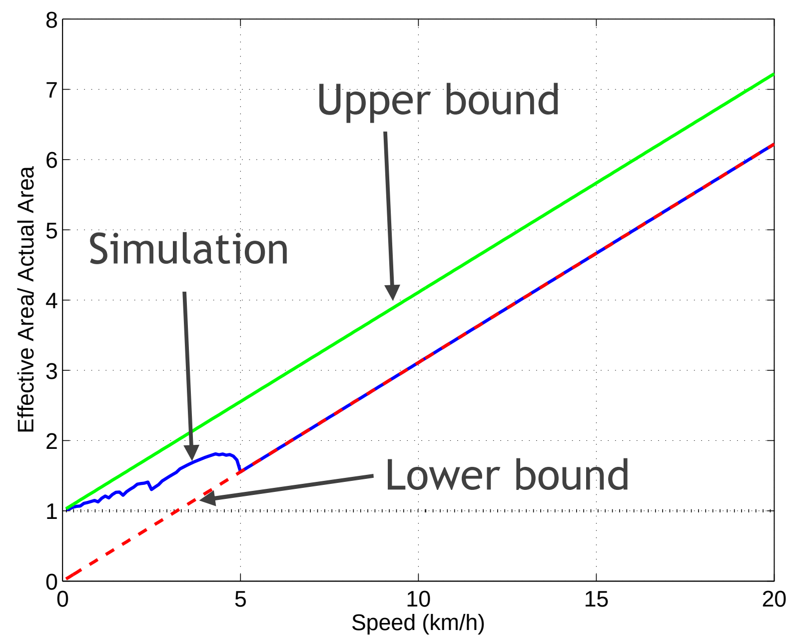
- When the network receives a packet but does not have a connection for a given MT with ISR activated, it buffers the packets and pages **both** overlay and underlay.
- with ISR deactivated, the network pages overlay or underlay.

Ghosting Effect for Overlay Paging with ISR



Upper and lower bounds on the relative increase in number of MTs registered with overlay due to ghosting is given by

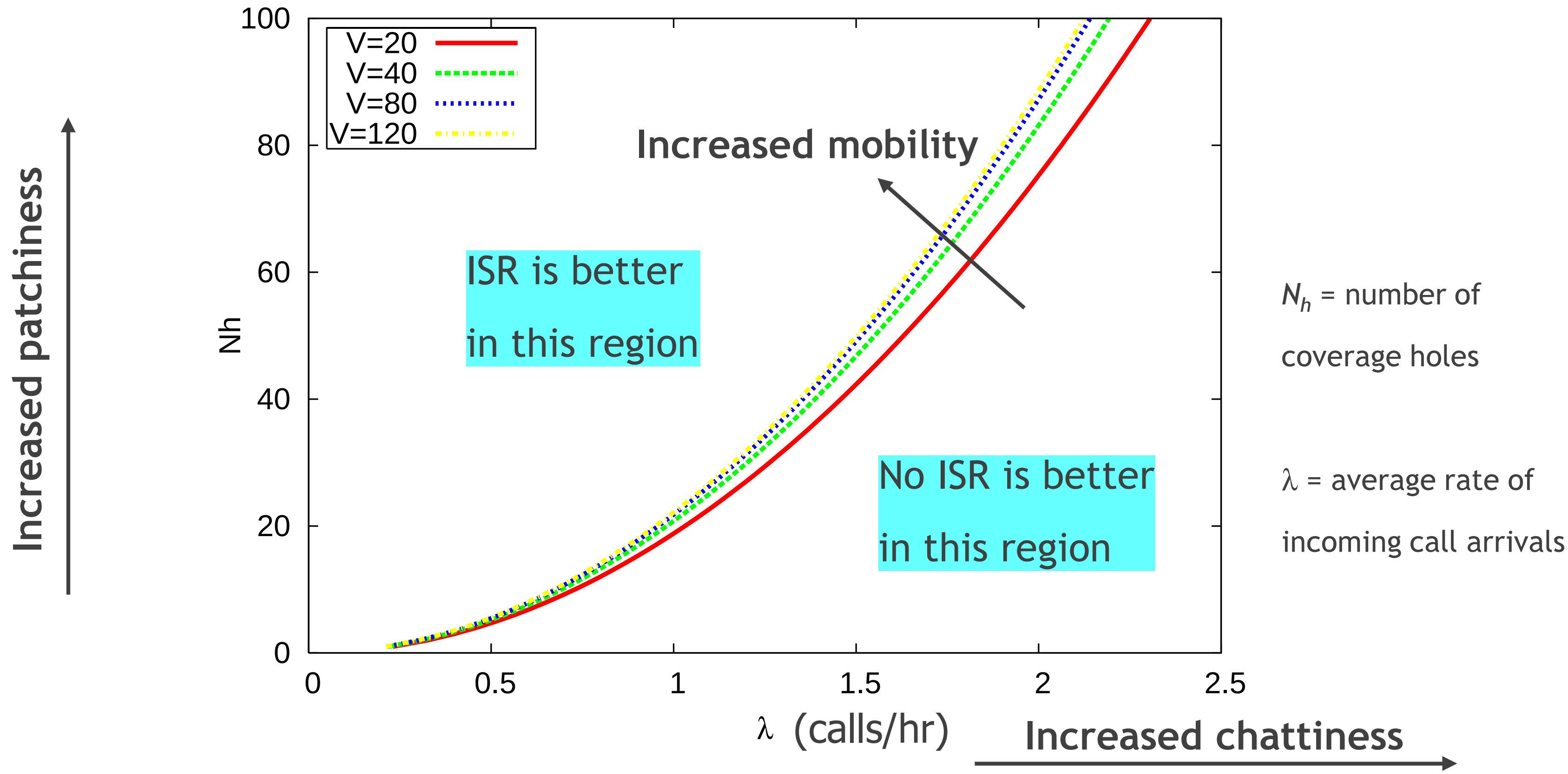
$$\max\left(1-\alpha, \frac{VT}{A_o/\bar{H}}\right) \leq \frac{N_s}{N_o} \leq 1 + \frac{VT}{A_o/\bar{H}}.$$



With ISR activated, ghosting increases paging load on overlay

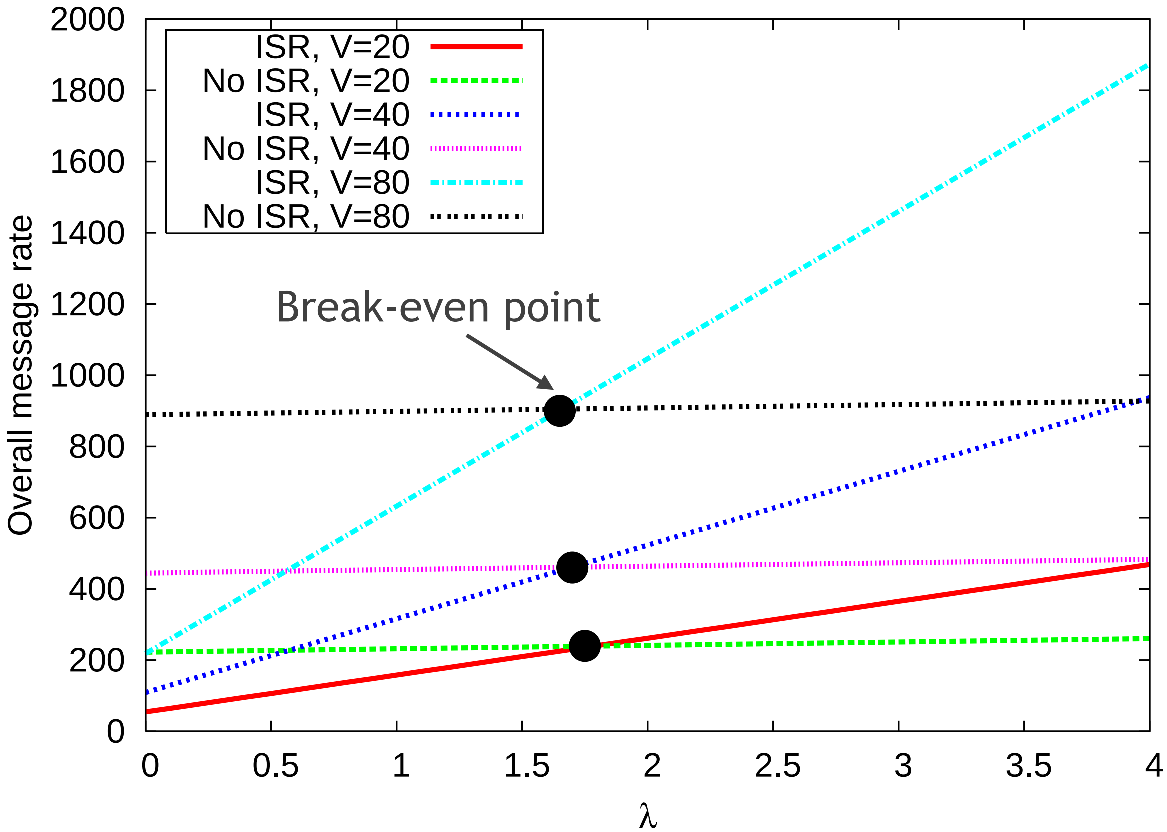
ISR is harmful

Tradeoff between Paging and Updating



- Break-even is when **Load of Updating-and-Paging *with ISR*** = **Load of Updating-and-Paging *without ISR***.
- Let $\lambda^*(V, N_h)$ be λ at break-even point.
– $\lambda^*(V, N_h)$ is sensitive to patchiness but insensitive to velocity.

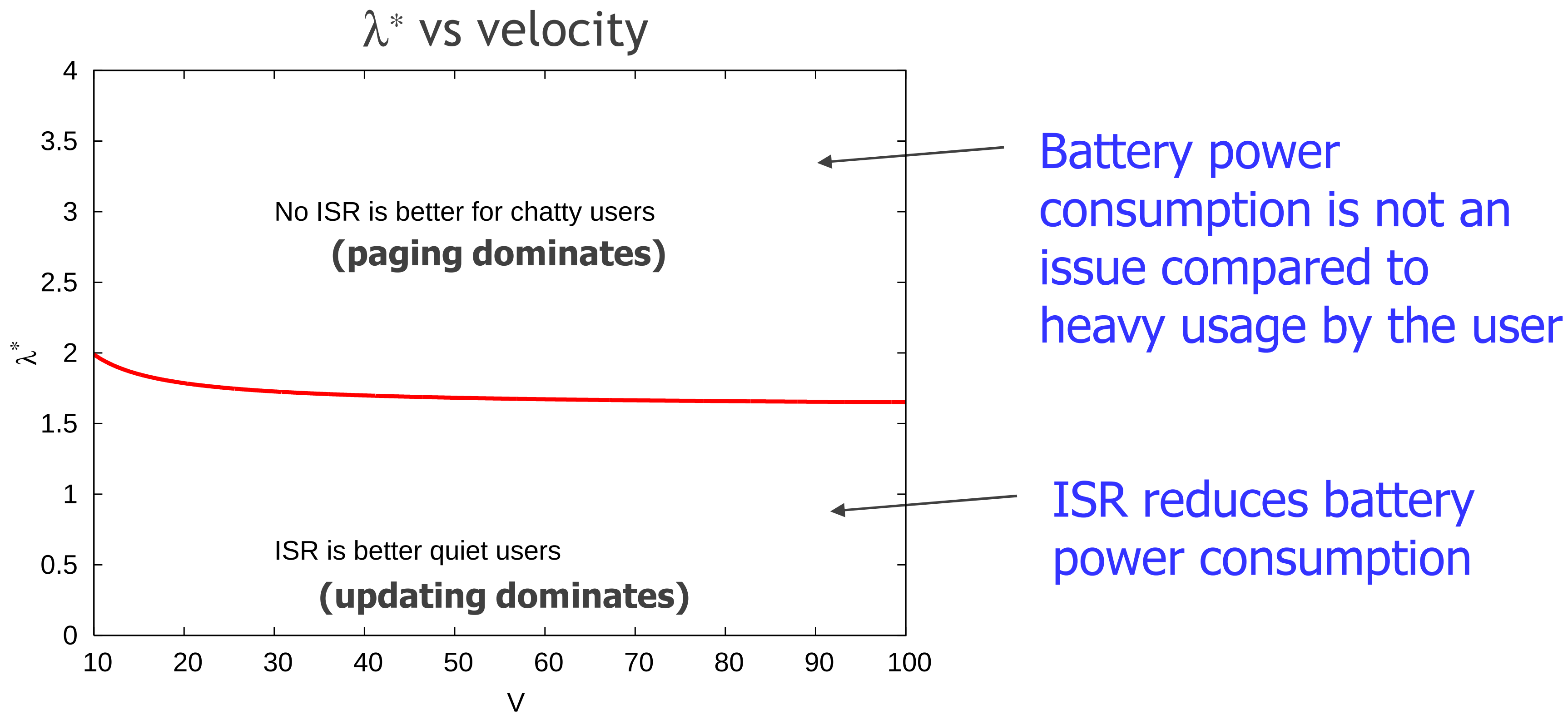
Signaling Load per MT for a Given Overlay



Signaling load vs λ for a given overlay deployment (N_h=60)

- With ISR activated, the signaling load depends strongly on λ, while with ISR deactivated, the load is insensitive to λ.
- Using ISR is beneficial when λ is below the break-even point but harmful above it.

Threshold-Based ISR



– $\lambda^*(V) = \lambda(V)$ at break-even point.

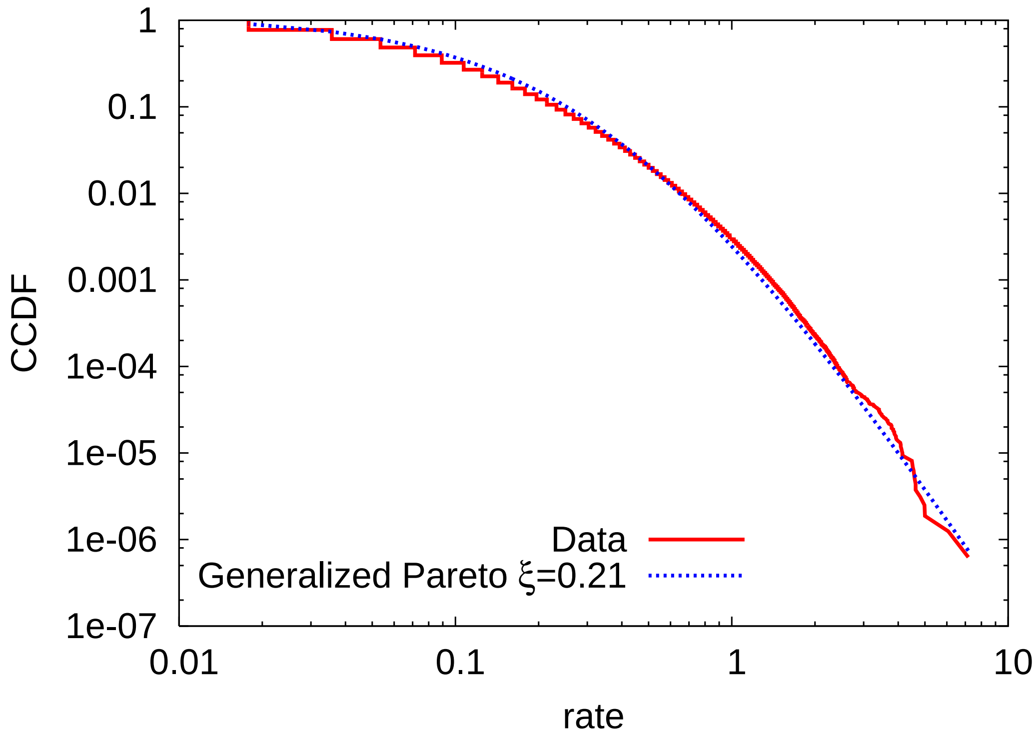
- Choose a *global* $\tilde{\lambda} = \lambda^*(V)$ at high-velocity value as high-velocity MT has more impact on load.
- This results in **threshold-based ISR** that is *independent* of MT velocity:
  - Activate ISR when $\lambda \leq \tilde{\lambda}$
  - Deactivate ISR when $\lambda > \tilde{\lambda}$

Experiment Setup

Distribution of call arrival rates of MTs follows a generalized Pareto:

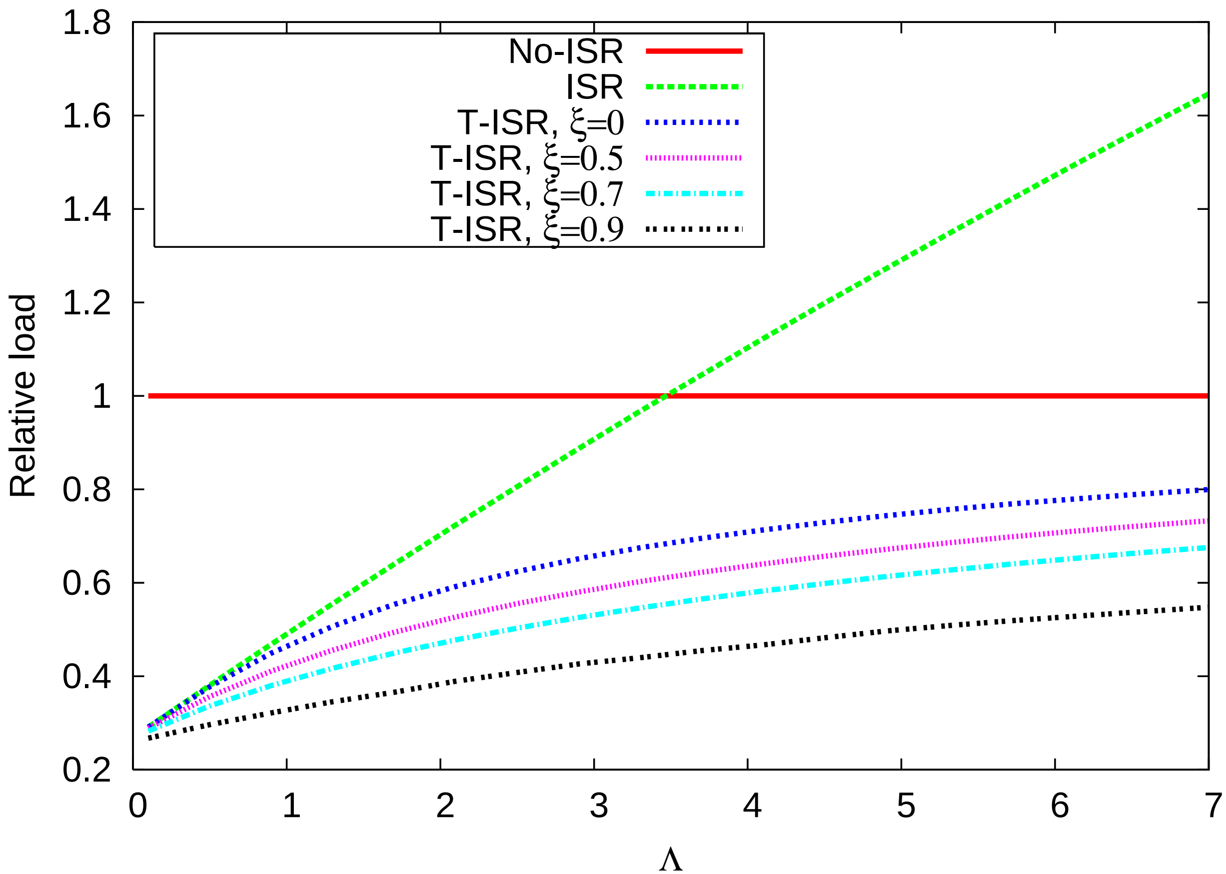
$$F(x;\xi,\sigma) = \begin{cases} 1 - (1 + \xi x/\sigma)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-x/\sigma), & \xi = 0, \end{cases}$$

where $\sigma = \Lambda\,(1-\xi)$, and Λ is aggregate call arrival rate per MT per hour.



Comparison between generalized Pareto and data from a real trace.

Result



Comparison of No-ISR, ISR and Threshold-based ISR.

Assumptions:
- Load normalized to no-ISR case.
- $\tilde{\lambda}$ = 1.7, Number of MTs = 500,000

Practical Setting of ISR

- Previous *open-loop* approach requires knowledge of overlay parameters (e.g., overlay size, number of holes, sizes and shapes of holes) to find the threshold $\tilde{\lambda}$ that minimizes signaling load, $M(\tilde{\lambda})$.
- In reality, the parameters of the overlay deployment are generally not known.

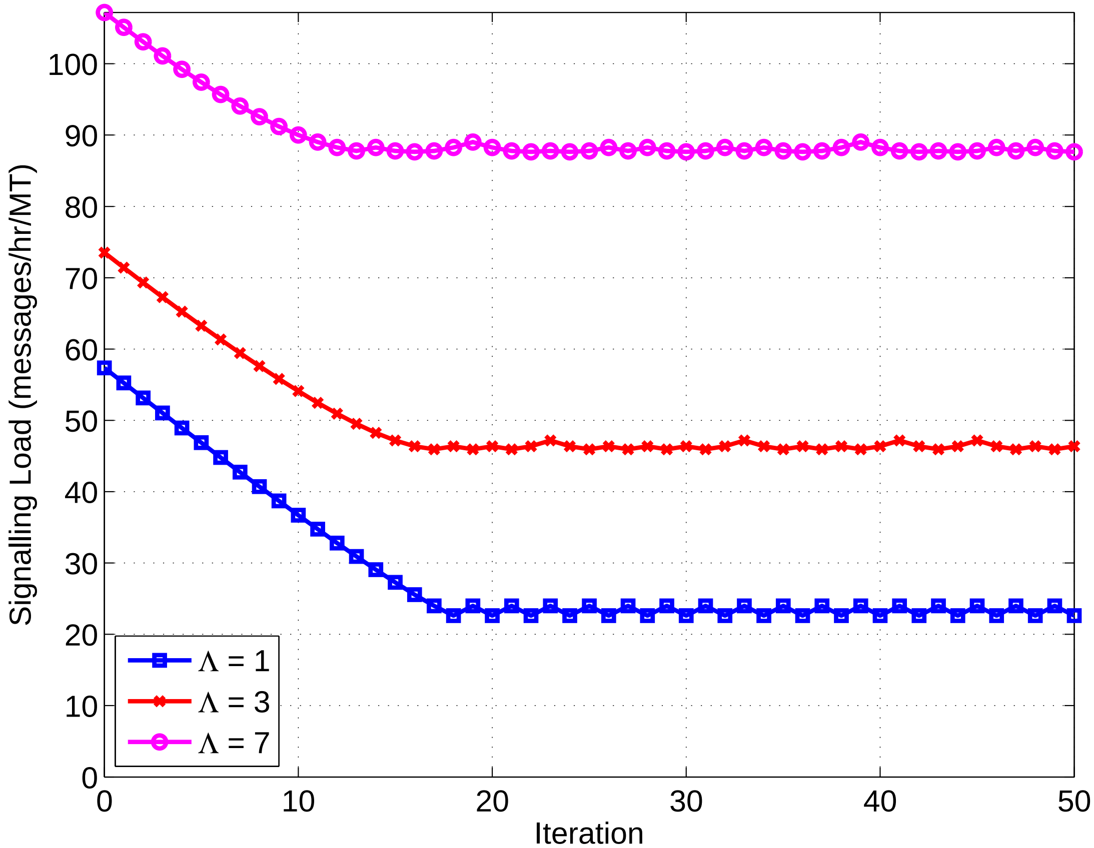
Measurement-based Approach

- An alternative *closed-loop* approach is to adopt a stochastic approximation algorithm (motivated by Kiefer-Wolfowitz algorithm) to iteratively optimize the threshold based on noisy observation $\hat{M}(\tilde{\lambda}, t_n, \tau)$ – an empirical estimate of measured signaling load taking into account random call arrivals in time interval $(t_n, t_n+\tau)$ with a control variable $\tilde{\lambda}$.
- Let $y=F(\tilde{\lambda})$ be the fraction of MTs with call arrival rate lower than $\tilde{\lambda}$ and let $q(y) = F^{-1}(y)$. Starting at $y_0 = 0$, the algorithm iteratively evaluates:

$$y_{n+1} = y_n + \beta \,\mathrm{sgn}\left(\hat{M}(q(y_n), t_n, \tau) - \hat{M}(q(y_n + \delta), t_n + \tau, \tau)\right)$$

increment step size (β) — perturbation step size (δ)

Result



- Progress of the stochastic approximation algorithm.
- After convergence, a single iteration can be run each day during the busy-hour period.

Assumptions:

- System with 10,000 MTs with individual $\{\lambda\}$ drawn from a generalized Pareto distribution with shape parameter ξ=0.7.
- δ=0.1, β=0.05, τ=30min

Conclusions

- The proliferation of multi-mode mobile terminals (MTs) can significantly stress signaling load in wireless networks.
- 3GPP has devised a mechanism to reduce signaling load called ISR, but no approach on how to set ISR is given or known.
- We analyze the tradeoff between updating and paging and quantify a single threshold value to decide on ISR activation.
- We develop a practical algorithm to activate or deactivate ISR for each MT without requiring knowledge of network deployment or terminal mobility.

Thanks!