Device-to-Device Load Balancing for Cellular Networks

Lei Deng, Ying Zhang, Minghua Chen, Jack Y. B. Lee, Ying Jun (Angela) Zhang

Zongpeng Li

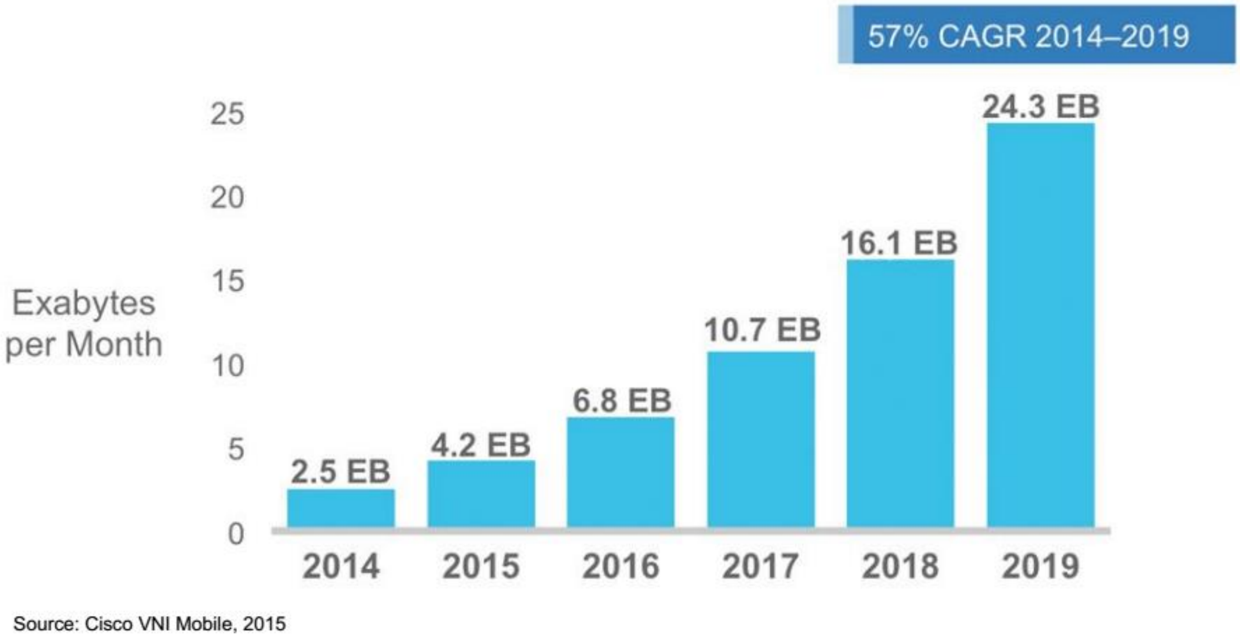
Lingyang Song

香港中文大學
The Chinese University of Hong Kong

UNIVERSITY OF CALGARY

北京大學
PEKING UNIVERSITY

Mobile Data Traffic Is Skyrocketing



Increased Mobile Data

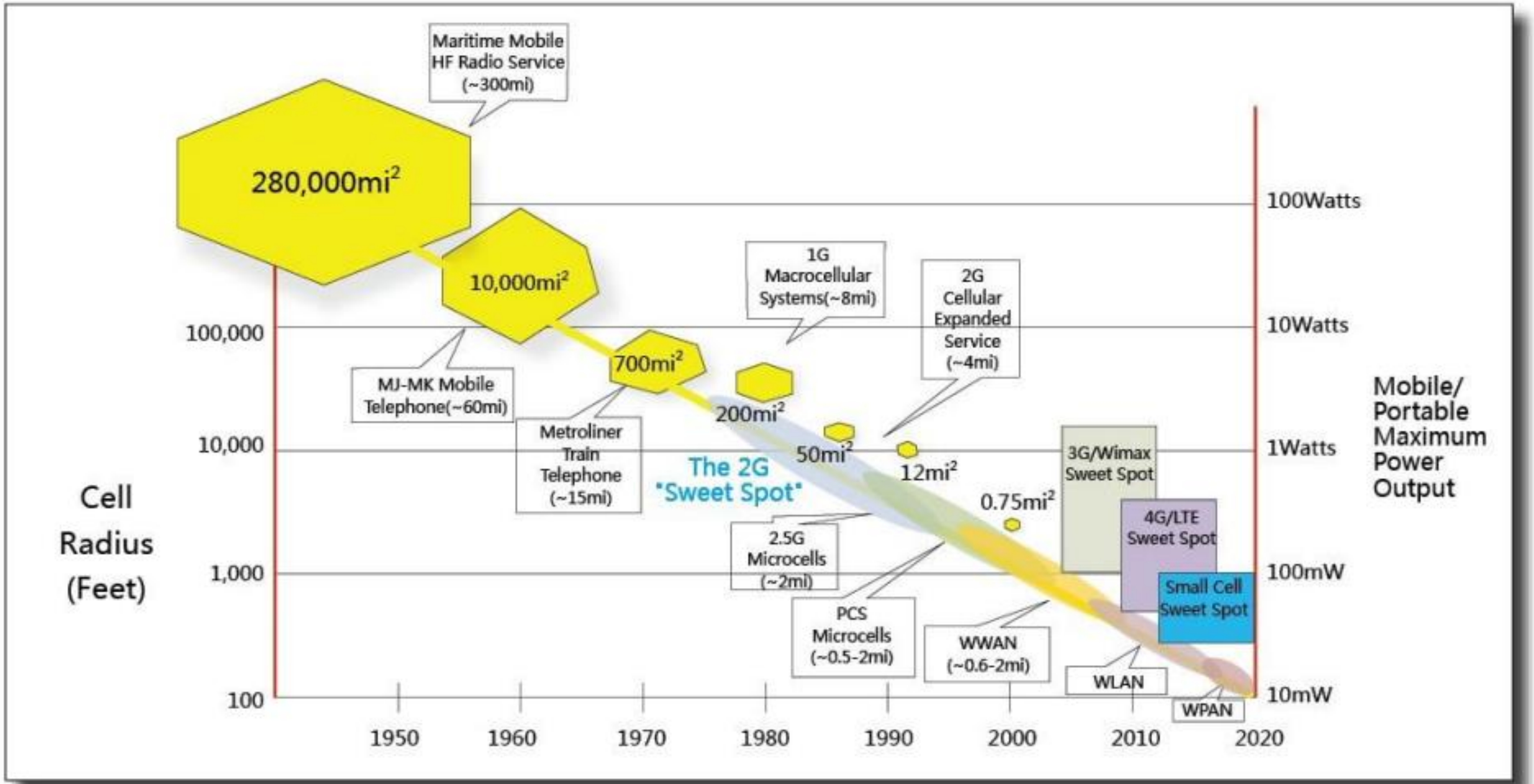
V.S.

Limited Radio Spectrum

- Cisco Forecasts 24.3 EB per Month of Mobile Data Traffic by 2019
- A **10x** Increase over 2014

24.3 EB (Exabyte) = **40%** of Monthly Global Fixed-Internet Traffic in 2014

The Cell Size Is Shrinking



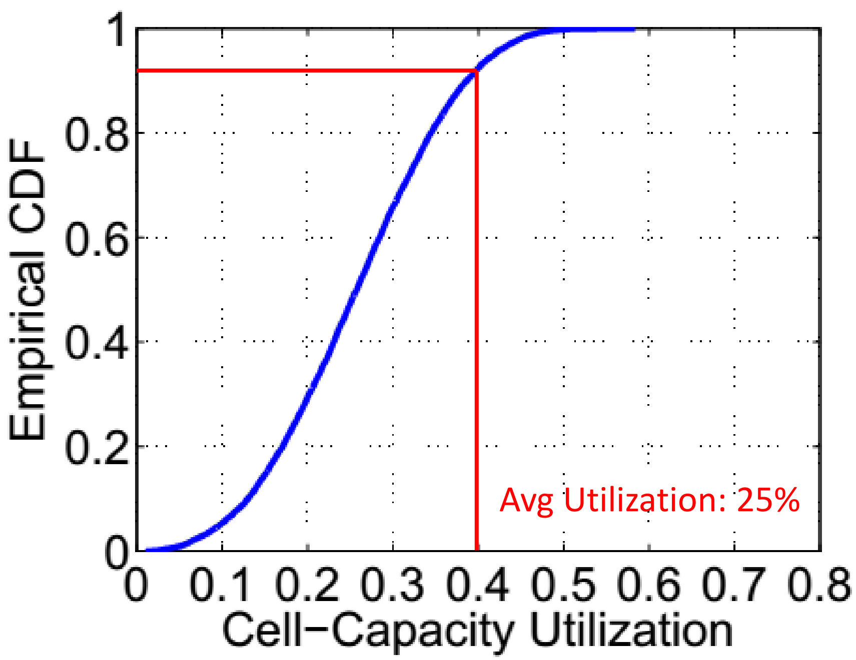
The Trend of Shrinking Cells (Source: ZTE Article)

Small Cell Improves Spectrum Spatial Efficiency

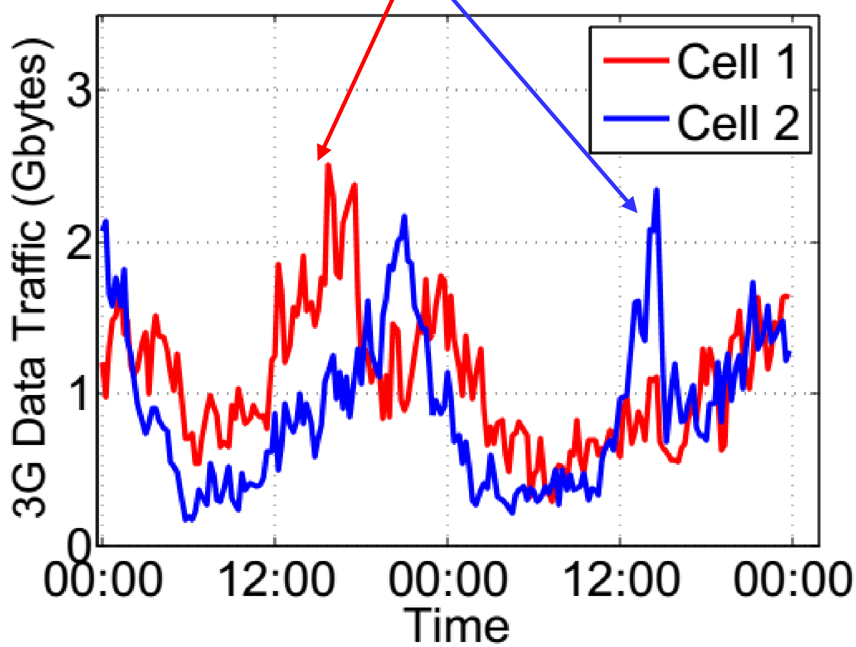
yet Degrades Spectrum Temporal Efficiency

Case Study: SmarTone

Average Cell Radius: 200m



Non-synchronized Peak Traffic

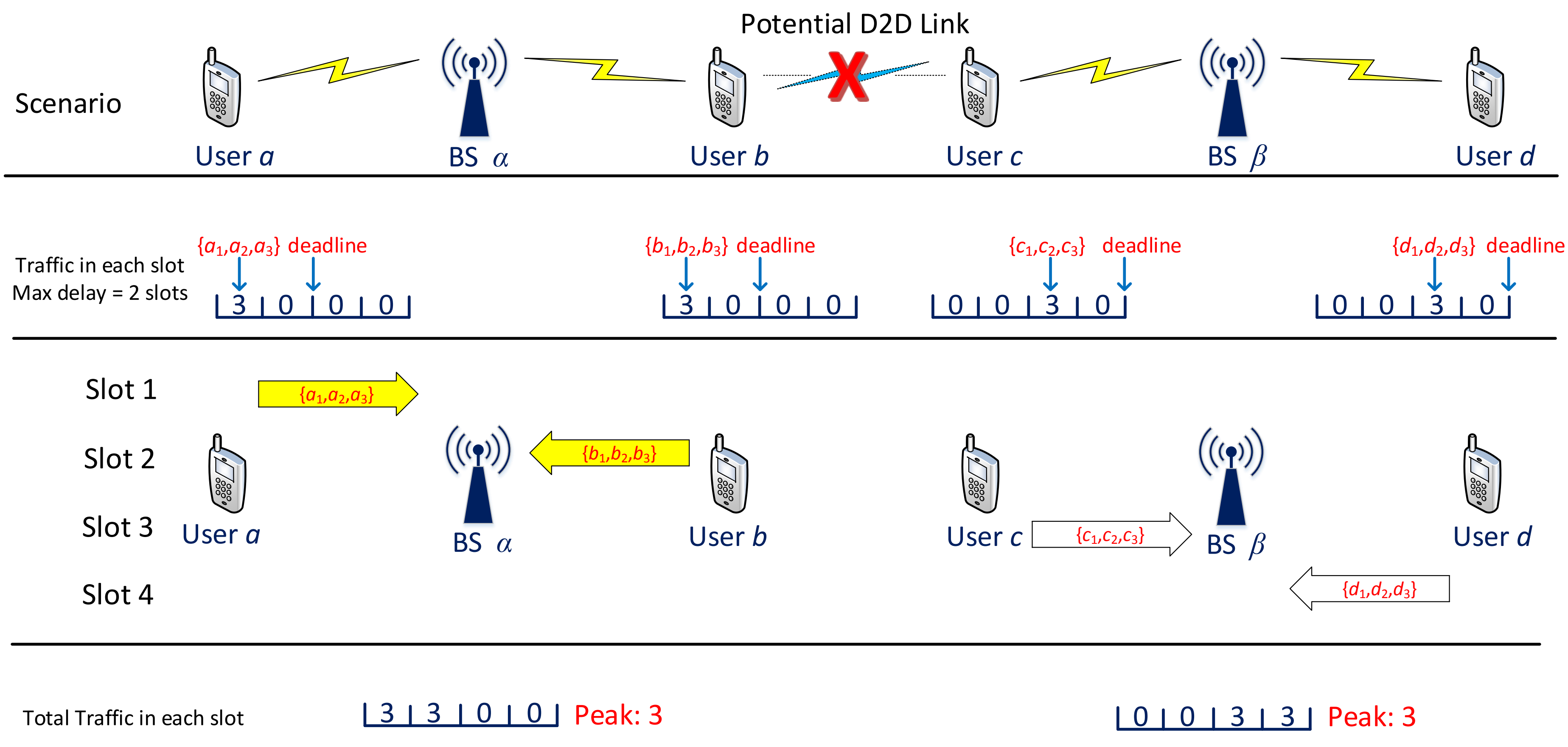


Low Spectrum Temporal Efficiency!

Load Balancing Can Potentially Increase Spectrum Temporal Efficiency

We Advocate ***Device-to-Device Load Balancing (D2D LB)*** Scheme

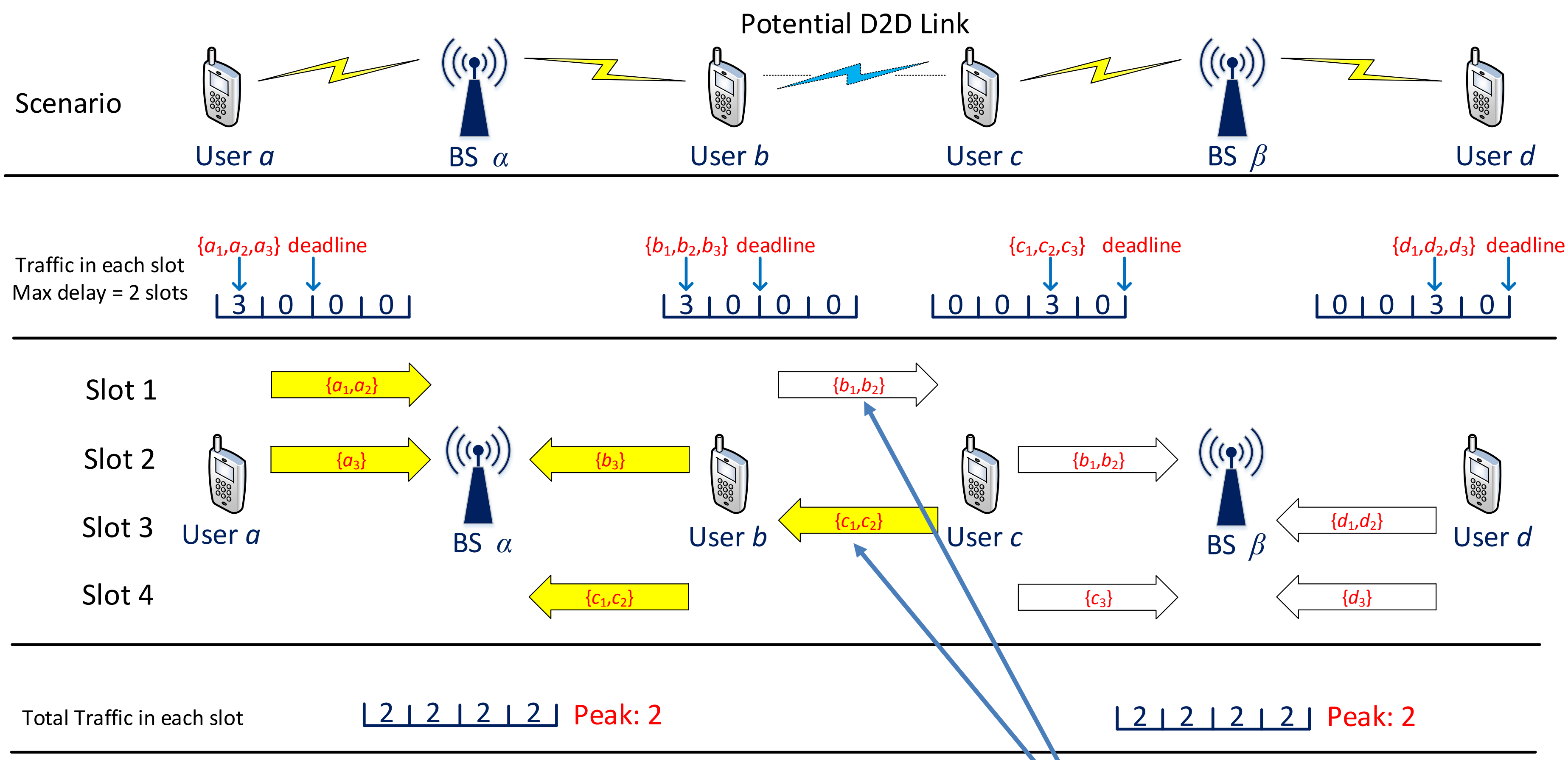
Example: Without D2D Load Balancing



Peak Traffic: 3

Spectrum Temporal Utilization: 50%

Example: With D2D Load Balancing



Peak Traffic: **3** → **2**

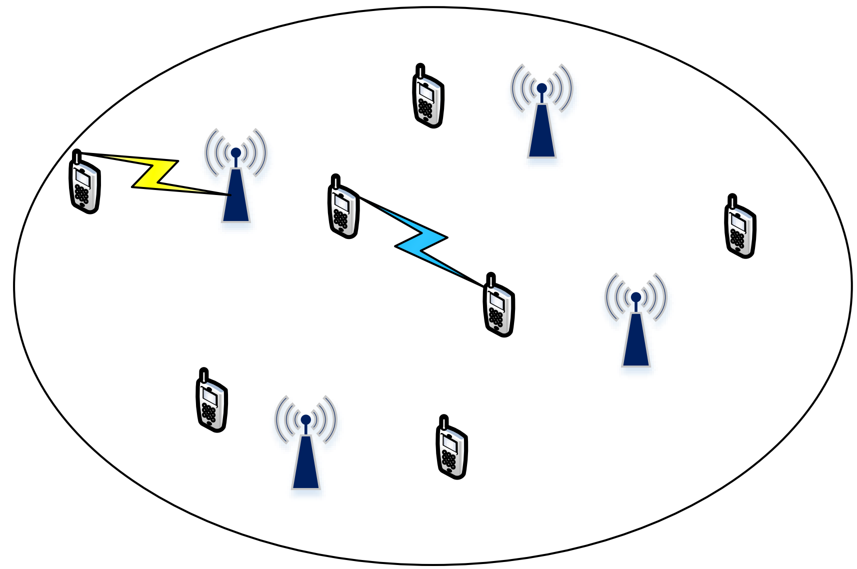
Spectrum Temporal Utilization: **50%** → **100%**

4 Extra D2D Transmissions

System Model

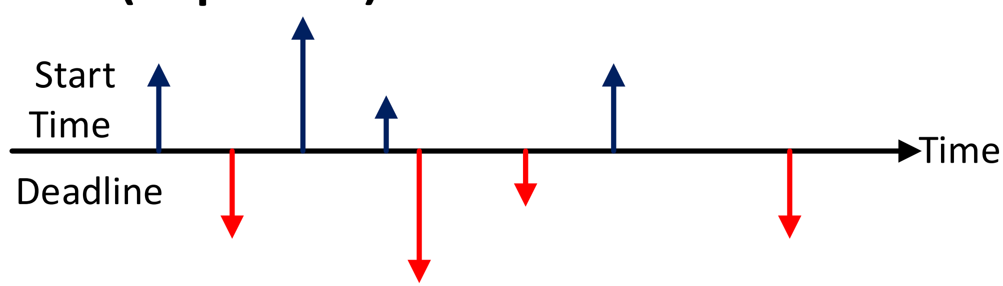
- Network topology
  - Directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
  - Link rate R_{uv}

    $$\text{Transmitted Volume per Slot} = \text{Link Rate} \times \text{Assigned Resources}$$



- Traffic demand pattern (Uplink)
  - $z^{s\tau} = (x^{s\tau}, d^{s\tau})$

    user start-time volume deadline



All traffic should reach **any** BSs before expiration!

Performance Metrics

- □ Sum peak traffic/resource reduction (Benefit)

$$\rho = \frac{P_{ND} - P_D}{P_{ND}} \in [0, 1)$$

  - P_{ND} is the minimal sum peak traffic without D2D
  - P_D is the minimal sum peak traffic with D2D LB

- □ D2D traffic overhead ratio (Cost)

$$\eta = \frac{V_{D2D}}{V_{D2D} + V_{BS}} \in [0, 1)$$

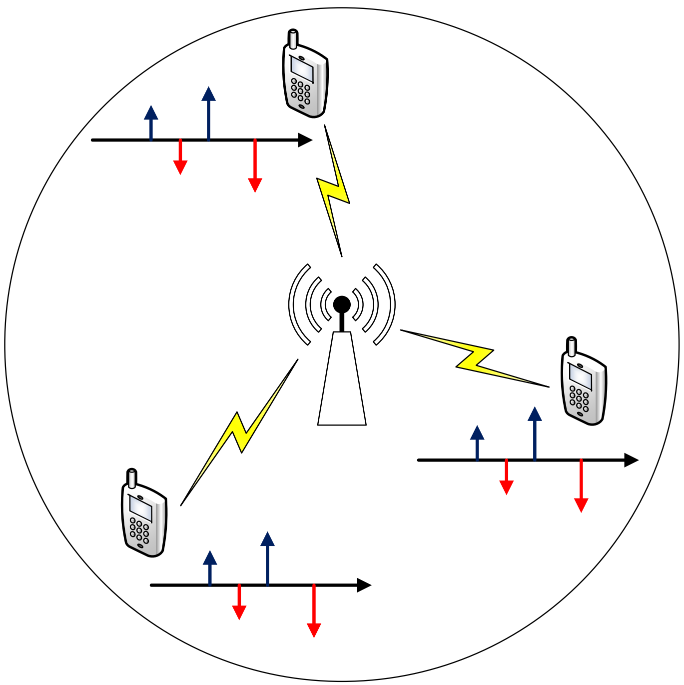
  - V_{D2D} is the sum volume of all D2D traffic
  - V_{BS} is the sum volume of all user-BS traffic

We optimize the benefit and characterize the corresponding cost

Minimize Sum Peak Traffic: No D2D

- We can use YDS algorithm to get the minimal peak traffic of any BS b
- Define the *intensity* of an interval $I = [z, z']$ as

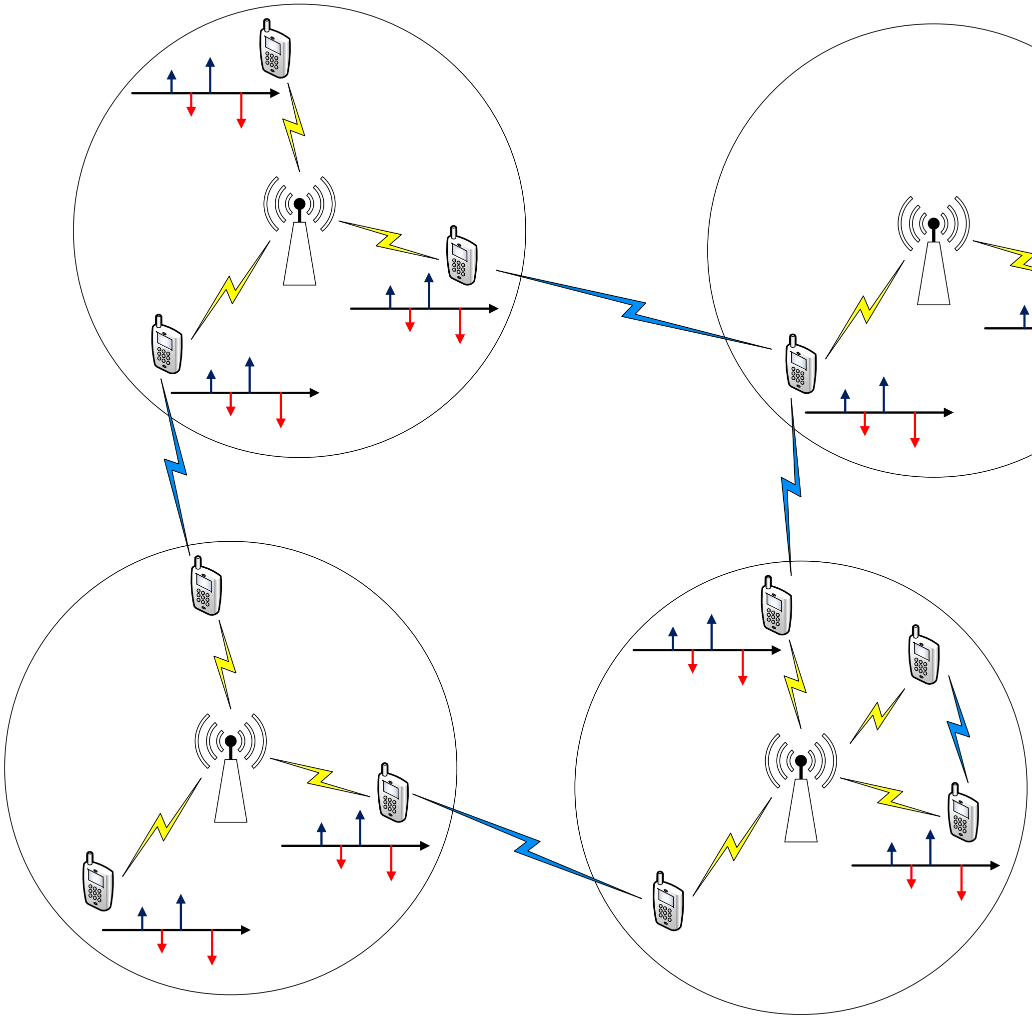
$$g_b(I) = \frac{\sum_{(s,\tau)\in\mathcal{A}_b(I)} \frac{x^{s\tau}}{R_{sb}}}{z' - z + 1}$$



- **Theorem**: $P_b^* = \max_{I\subset[1,T]} g_b(I).$

F. Yao, A. Demers and S. Shenker, "A scheduling model for reduced CPU energy," in *Proc IEEE FOCS*, 1995.

Minimize Sum Peak Traffic: D2D LB



All BSs Are Coupled!

A Large-Scale LP

No Efficient Algorithm Now

Limitations of Conceivable Approach

- No closed-form expression
  - Minimal sum peak traffic with/without D2D LB
  - Sum peak traffic reduction
- No efficient algorithm
  - Minimal sum peak traffic with D2D LB
- Hard to get **insights** of the benefit of D2D LB

Sum Peak Traffic Reduction: Upper Bound

- **Theorem**: For an arbitrary network topology and an arbitrary traffic pattern,

$$\rho = \frac{P_{ND} - P_D}{P_{ND}} \leq \frac{\max\{r, 1\} + \tilde{r}\Delta^- - 1}{\max\{r, 1\} + \tilde{r}\Delta^-}.$$

Captures the link-rate advantages of **intra-cell D2D links** over the **user-BS links**

Captures the link-rate advantages of **inter-cell D2D links** over the **user-BS links**

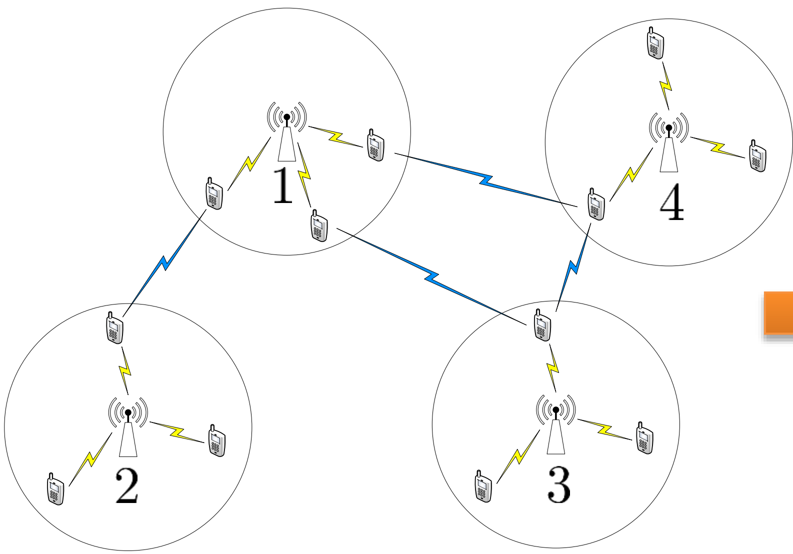
Captures the BS-level network connectivity and **traffic aggregation** capability

Sum Peak Traffic Reduction: Upper Bound

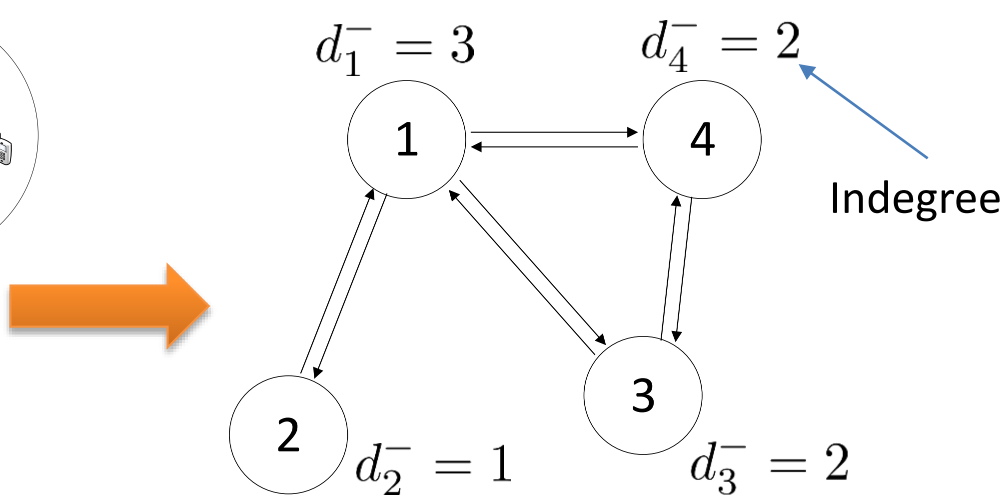
- **Corollary**: If $r = \tilde{r} = 1$, then we have

$$\rho = \frac{P_{ND} - P_D}{P_{ND}} \leq \frac{\Delta^-}{\Delta^- + 1}$$

Network Topology



D2D Communication Graph (BS-level)



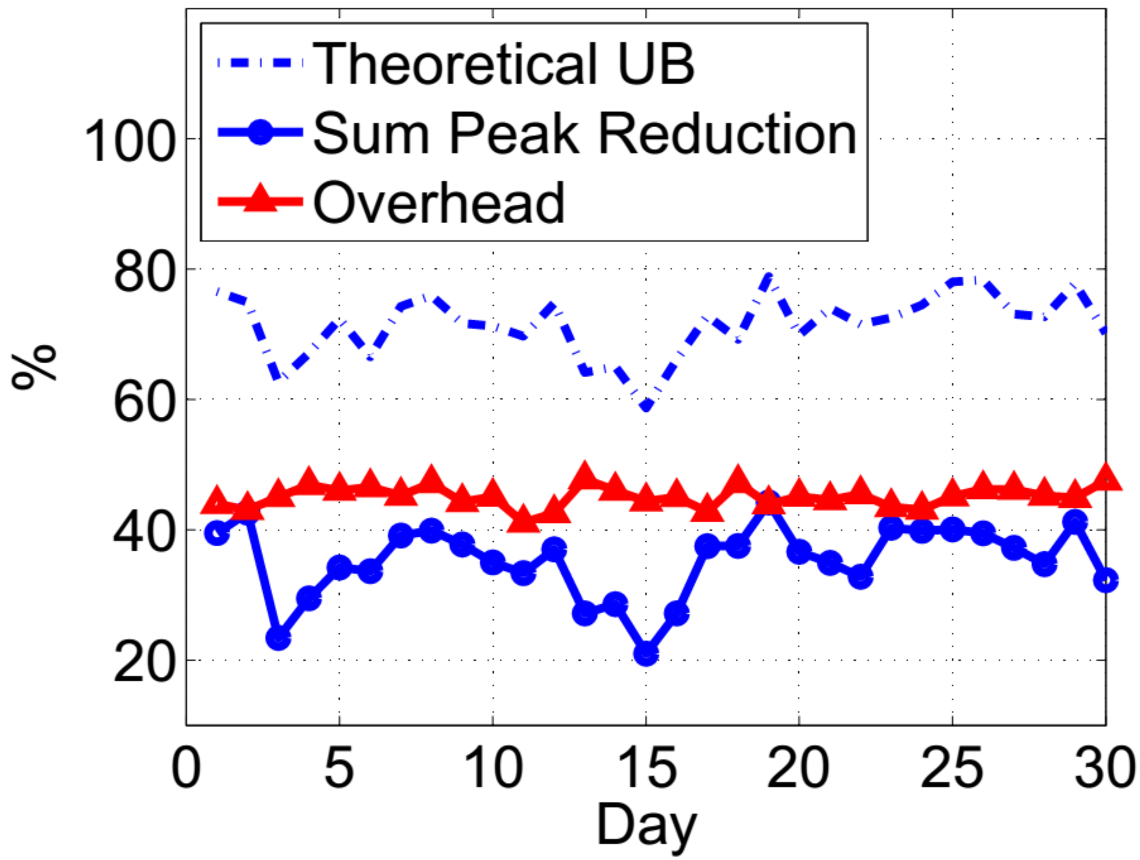
Max Indegree: $\Delta^- = \max_i d_i^- = 3$

$$\rho \leq \frac{\Delta^-}{\Delta^- + 1} = 75\%$$

Discussions

- **Corollary**: $\rho = \frac{P_{ND} - P_D}{P_{ND}} \leq \frac{\Delta^-}{\Delta^- + 1}$
- Δ^- evaluates the traffic aggregation capability
- The more traffic each BS aggregates for other BSs, the more statistical multiplexing gain
- How good is this upper bound?
  - $\rho \to \frac{2}{3} = \frac{\Delta^-}{\Delta^- + 1}$ in the ring topology
  - i.e., tight under ring topology ($\Delta^- = 2$)

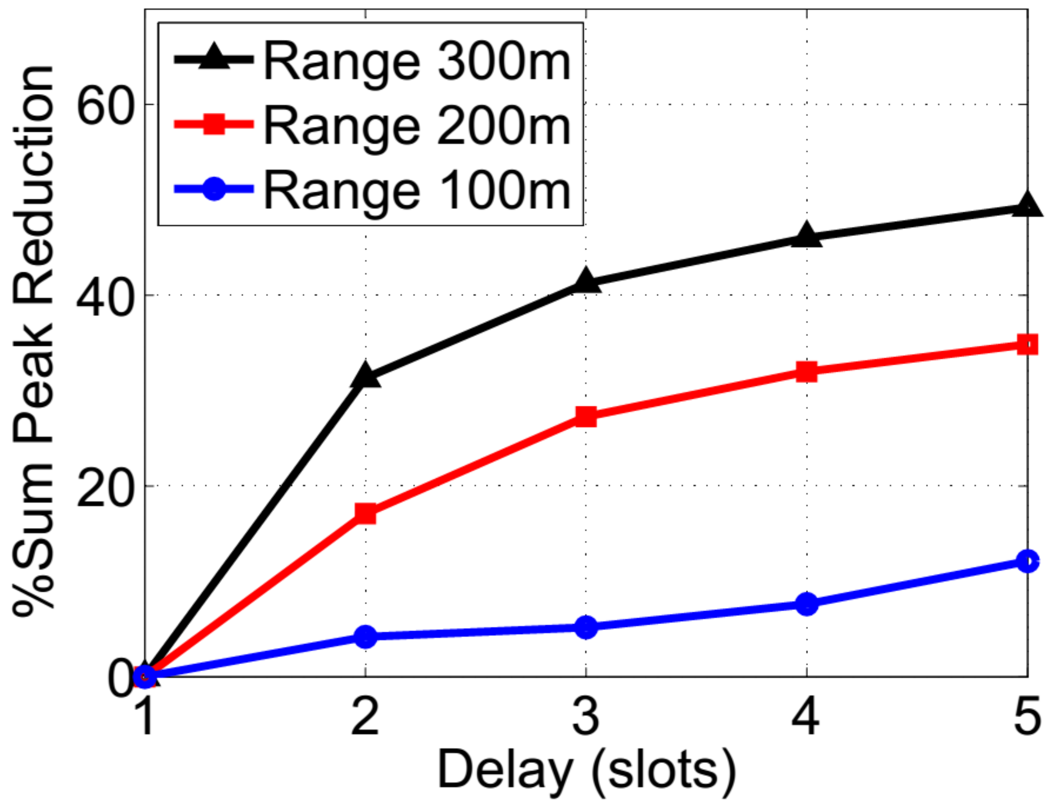
Trace-driven Simulation: Benefit and Cost



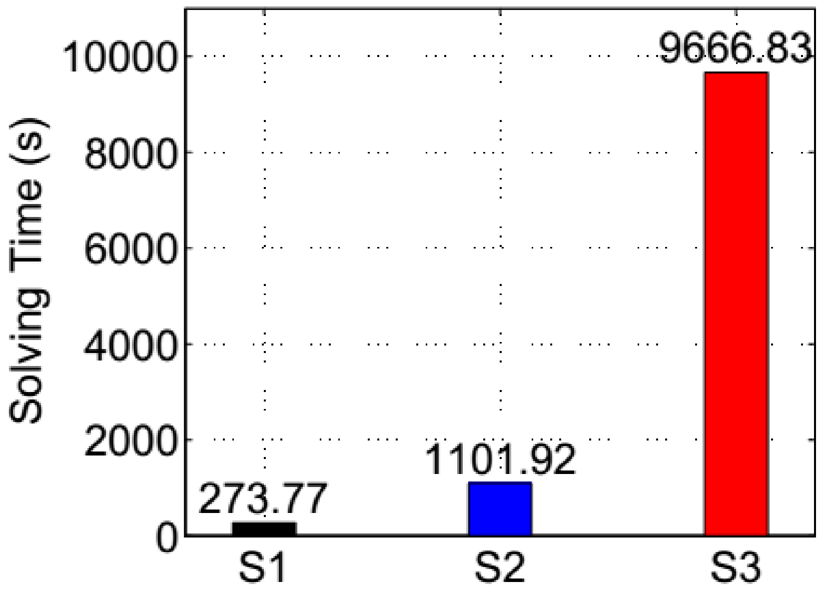
Benefit: 35%

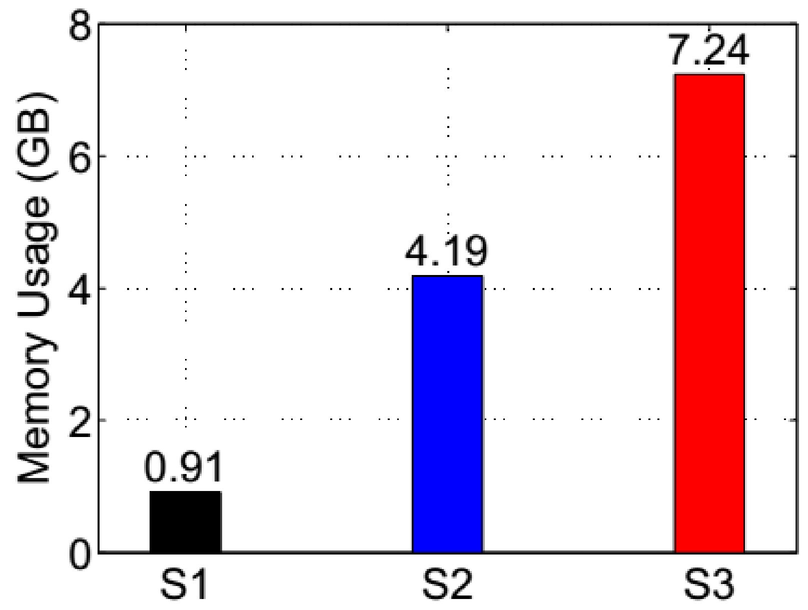
Cost: 45%

Effects of Traffic Delay and Commu. Range



Computational Cost of Large-Scale LP





Instance	$\lvert\mathcal{B}\rvert$	$\lvert\mathcal{U}\rvert$	$\lvert\mathcal{E}\rvert$	# of demands	T
S1 (Light)	3	15	139	4035	43200
S2 (Medium)	6	30	344	6945	43200
S3 (Heavy)	9	45	1083	10095	43200

Conclusion

- Advocate the concept of D2D load balancing
- Define the performance metrics for both benefit and cost
- Theoretical upper bound for arbitrary settings
- Real-world trace-driven simulations

Future Work

- Design efficient algorithms for sum peak traffic minimization with D2D LB
- Design incentive mechanisms for D2D users
- Distributed/Online scheduling algorithms
- Refine the physical-layer channel model and relax some assumptions

Q&A

Thank you!

Backup Slides

Minimize Sum Peak Traffic: No D2D

$$
\begin{aligned}
\min \quad & P_b \\
\text{s.t.} \quad & \sum_{t=\tau}^{d^{s\tau}} y_{sb}^{s\tau}(t) R_{sb} = x^{s\tau}, \forall s \in \mathcal{U}_b, \tau \in [1, T] \\
& \sum_{s \in \mathcal{U}_b} \sum_{\tau : \tau \le t \le d^{s\tau}} y_{sb}^{s\tau}(t) = \alpha_b(t), \forall t \in [1, T] \\
& \alpha_b(t) \le P_b, \forall t \in [1, T] \\
& y_{sb}^{s\tau}(t) \ge 0, \forall s \in \mathcal{U}_b, \tau \in [1, T], t \in [\tau, d^{s\tau}] \\
\text{var} \quad & y_{sb}^{s\tau}(t), \alpha_b(t), P_b
\end{aligned}
$$

Minimize Sum Peak Traffic: D2D

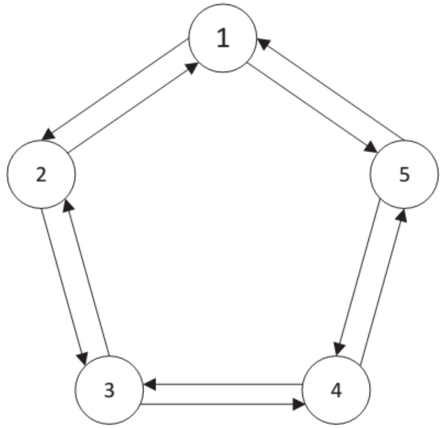
$$
\begin{aligned}
\min \quad & \sum_{b \in \mathcal{B}} P_b \\
\text{s.t.} \quad & \text{feasible traffic scheduling policy}, \\
& \sum_{v \in \mathcal{U}_b} \sum_{s \in \mathcal{U}} \sum_{\tau : \tau \le t \le d^{s\tau}} y_{vb}^{s\tau}(t) = \alpha_b(t), \\
& \qquad \qquad \forall b \in \mathcal{B}, t \in [1, T] \\
& \sum_{u \in \mathcal{U}_b} \sum_{v \in \mathrm{in}(u) \backslash \{u\}} \sum_{s \in \mathcal{U}} \sum_{\tau : \tau \le t \le d^{s\tau}} y_{vu}^{s\tau}(t) = \beta_b(t), \\
& \qquad \qquad \forall b \in \mathcal{B}, t \in [1, T] \\
& \alpha_b(t) + \beta_b(t) \le P_b, \forall b \in \mathcal{B}, t \in [1, T] \\
\text{var} \quad & y_{uv}^{s\tau}(t), \alpha_b(t), \beta_b(t), P_b
\end{aligned}
$$

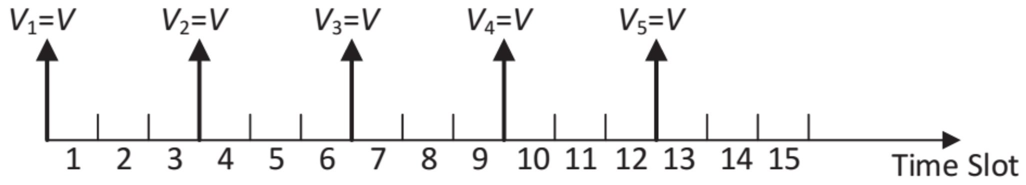
Ring Topology

- For any $D \geq 1$, there exists a ring topology and a traffic demand pattern such that

$$\rho = \frac{2(D-1)}{3D-2}$$

  - $\lim_{D\to\infty} \rho = \frac{2}{3} = \frac{\Delta^-}{\Delta^-+1}$
  - The bound is **asymptotically tight**





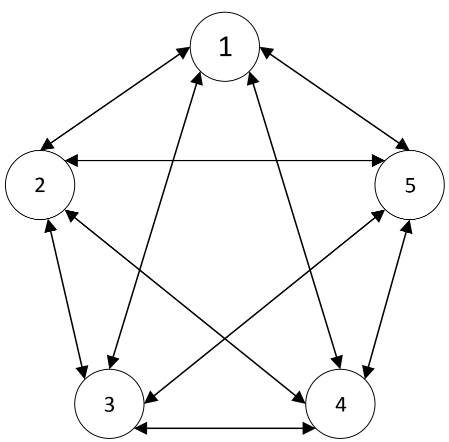
$(D = 3)$

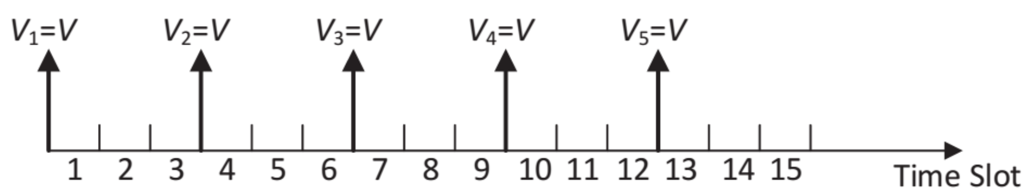
Complete Topology

- In a *N*-BS complete topology, there exists a traffic demand pattern such that

$$\rho = \frac{N-1}{N+1}$$

  - $\lim_{N\to\infty} \rho = 1$
  - In ***the best case***, we can achieve **100%** sum peak traffic reduction!





$$(N = 5)$$

Tradeoff between Benefit and Cost

- Tradeoff between sum peak traffic reduction ρ and overhead ratio η