AFRL HERITAGE | 1917–2017

100 YEARS OF U.S. AIR FORCE SCIENCE & TECHNOLOGY

A Multi-Agent Q-Learning Based Rendezvous Strategy for Cognitive Radios

27 Jun 2017

Clifton Watson

Air Force Research Laboratory

Integrity ★ Service ★ Excellence

Outline

- Introduction
- Blind Rendezvous Problem
- Existing Approaches
- Proposed Method - A MAQLR Strategy
- System Model & Assumptions
- Application of Q-Learning to Blind Rendezvous
- MAQLR Strategy - Description
- Simulation Results
- Conclusion

Introduction

- **Public and private sectors rely on spectrum access.**
- **Increasing demands require efficient spectrum use.**
- **This can be provided by cognitive radios (CRs) that can**
  - **sense, learn, and adapt to spectrum**
  - **access unused/underused licensed spectrum as unlicensed secondary users (SUs)**

Blind Rendezvous Problem

- **SUs must quickly find each other to communicate in multi-channel spectrum environment.**
  - **No dedicated control channel or central entity can be used.**
  - **All channels can be used for rendezvous and data exchange.**
- **Challenges**
  - **PU and SU activities are random and unpredictable.**
    - **Minimize PU interference (PUI)**
    - **Avoid SU collisions**

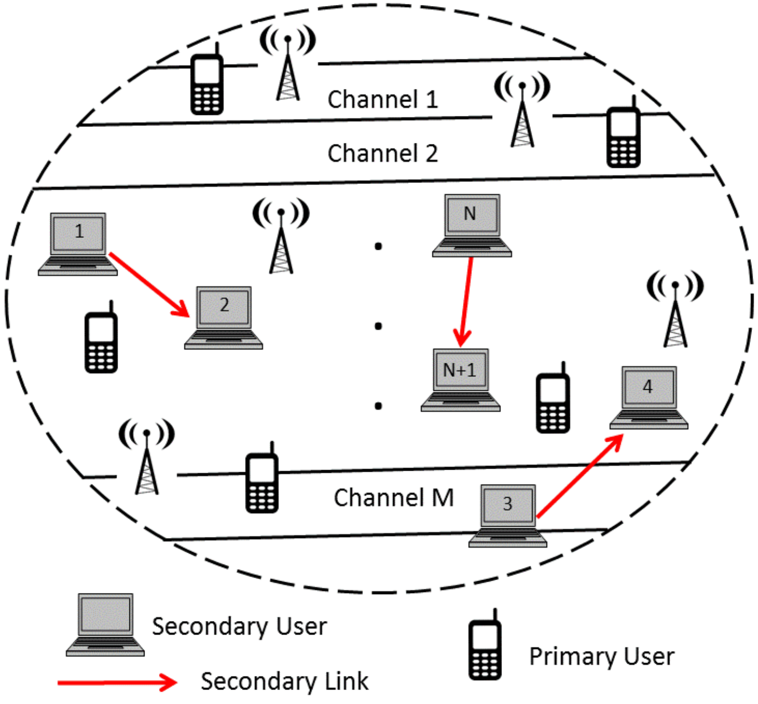
Existing Approaches

- Channel hopping (CH) is most common approach.
  - Predetermined CH sequences
    - Not biased towards any channels
    - Vulnerable to PUI and collisions
  - Adaptive CH sequences
    - Biased towards channels with least detected PU activity
    - Robust to PUI but more vulnerable to SU collisions

Proposed Method: A MAQLR Strategy

- **Multi-Agent Q-Learning Rendezvous (MAQLR) Strategy**
  - **SUs actively learn which channels are best for rendezvous.**
  - **Learning is based on exploration of spectrum environment.**
  - **Learned channels are generally less prone to PUI and SU collisions.**

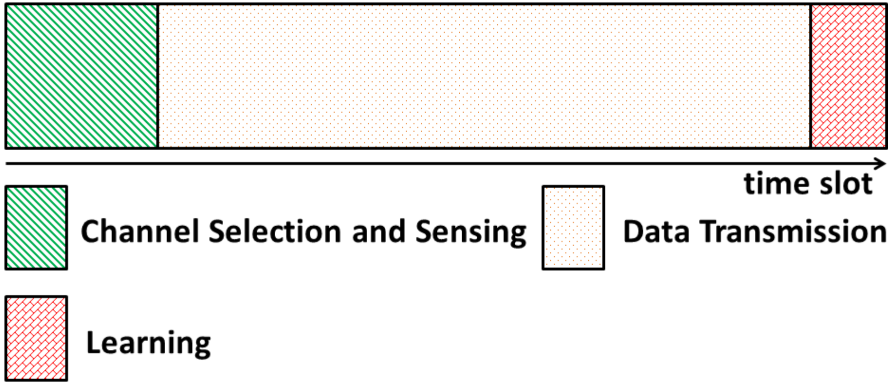
System Model & Assumptions



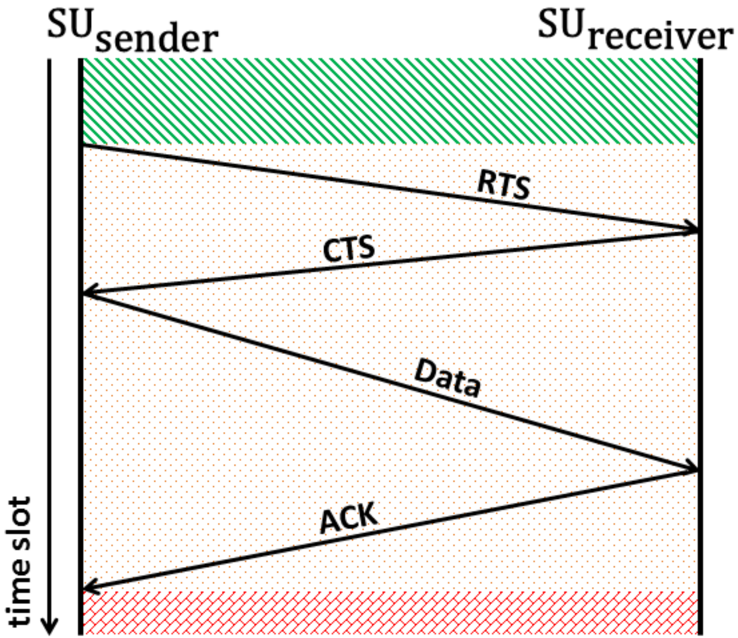
- N SU pairs (sender and receiver SU)
- M licensed channels ($1 \leq m \leq M$)
- Localized channel availability (θ_m)
- Slotted channel access by PUs and SUs
- Rendezvous in single slot with RTS/CTS

- SU Assumptions
  - start rendezvous at same time
  - sense the same PU activity
  - can distinguish between PU and SU
  - can access one channel at a time
  - do not exchange info with each other

System Model & Assumptions (cont'd)



Transmission structure for SUs



RTS/CTS Handshake

- SUs sense correctly with probability s.
- Sender SUs transmit RTS with probability p.

Application of Q-Learning to Blind Rendezvous (Cont'd)

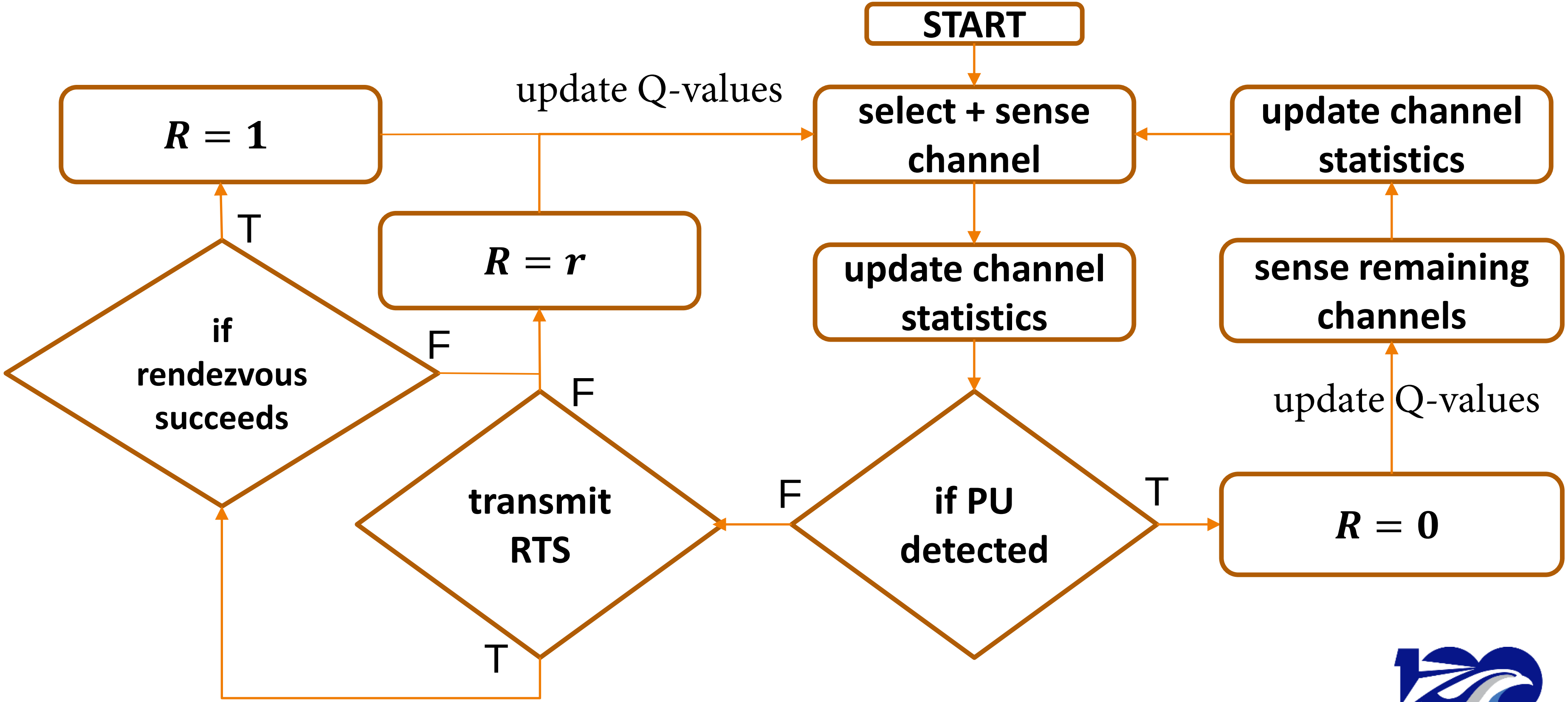
- Reward Strategy
  - Reward received for channel m at end of slot j
  - $R_m(j) = \begin{cases} 0, \text{PU activity is detected} \\ r, \text{PU activity is not detected and rendezvous fails} \\ 1, \text{PU activity is not detected and rendezvous succeeds} \end{cases}$
  - r is a random reward, $(0 < r < 1)$ *
- Rendezvous fails when
  - RTS not transmitted
  - Paired SUs select different channels
  - Collision occurs
  - Poor channel quality
- Rewards capture dynamics of PU and SU activity, as well as channel quality.

* For sake of brevity, audience is referred to paper for computation of r.

Application of Q-Learning to Blind Rendezvous (Cont'd)

- **Q-values updated for channel m at end of slot j**
  - $Q_m(j+1) = \left(1 - \alpha_m(j)\right)Q_m(j) + \alpha_m(j)R_m(j)$
  - $\alpha_m(j)$ **is the learning rate,** $(0 \leq \alpha_m(j) \leq 1)$
    - **determines how much old info is valued over new info**
    - **starts at 1 and decreases over time**
- **Probability of selecting channel m in slot j**
  - $P_m(j) = \dfrac{e^{Q_m(j)/W}}{\sum_m e^{Q_m(j)/W}}$
  - **balances tradeoff between exploration and exploitation**
  - W **is temperature parameter**
    - **decreases exploration over time**
    - **set to $\alpha_m(j)$ to decrease exploration by learning rate**

MAQLR Strategy - Description



Simulation Results

- **Simulation Setup**
  - **5-channel spectrum environment with localized availability**
  - $\theta_1 = 0.5, \theta_2 = 0.3, \theta_3 = 0.6, \theta_4 = 0.4, \theta_5 = 0.7$
  - $s = 0.9$
  - p varies: $0.7, 0.8, 0.9, 1$
  - **Number of SU pairs vary from 1 to 10.**
  - **Sender SUs assumed to always have data to send.**
  - **Compared against existing adaptive techniques**
    - **Enhanced Adaptive Multiple Rendezvous Control Channel with Variable Slots (EAMRCC-VS)**
    - **Nested Grid Quorum Frequency Hopping (NGQFH)**
    - **Follow same sensing procedure as MAQLR strategy**

Simulation Results (Cont'd)

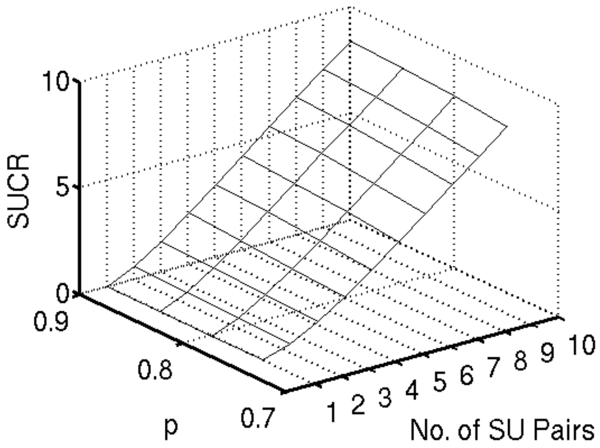
- Numbers of SU pairs on learned channels ($s = 0.9, p = 0.9$)

No.	$\theta_1 = 0.5$	$\theta_2 = 0.3$	$\theta_3 = 0.6$	$\theta_4 = 0.4$	$\theta_5 = 0.7$
1	0	0	0	0	1
2	0	0	1	0	1
3	1	0	1	0	1
4	1	0	1	1	1
7	1	1	2	1	2
8	2	1	2	1	2
9	2	1	2	2	2

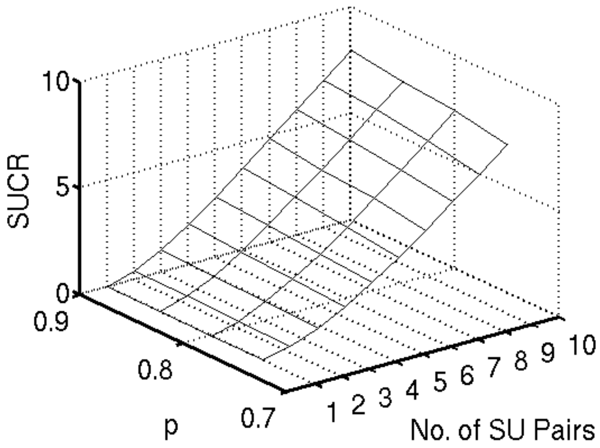
- SUs learn to use channels in effective and efficient manner.

Simulation Results (Cont'd)

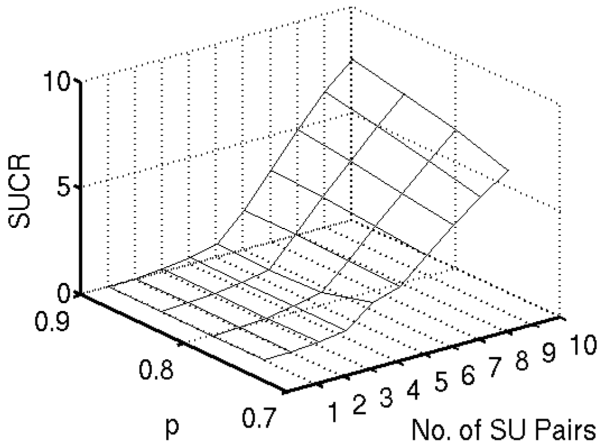
- **SU Collision Rate (SUCR)**
  - **Average number of SU collisions per RTS transmission**



EAMRCC-VS



NGQFH

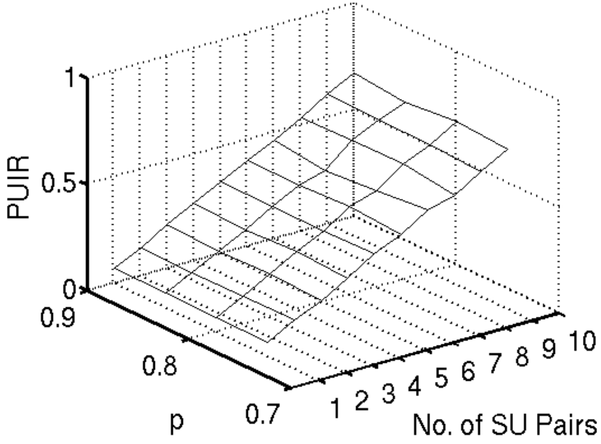


MAQLR

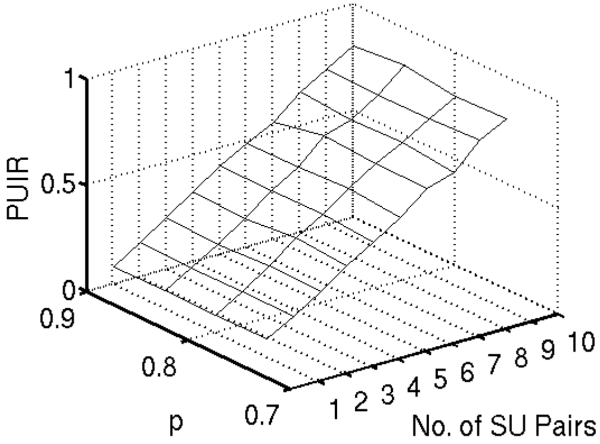
- SU's use of channels cause MAQLR strategy to have lower SUCR.

Simulation Results (Cont’d)

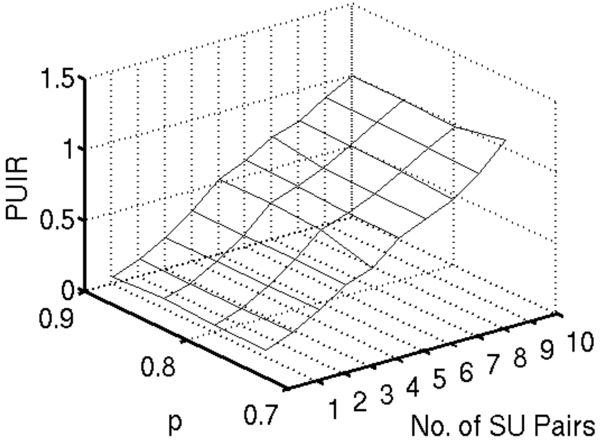
- **PUI Rate**
  - **Average number of PUIs per RTS transmission**



EAMRCC-VS



NGQFH

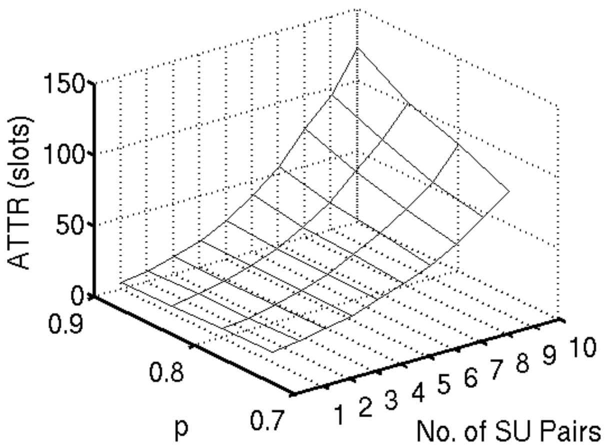


MAQLR

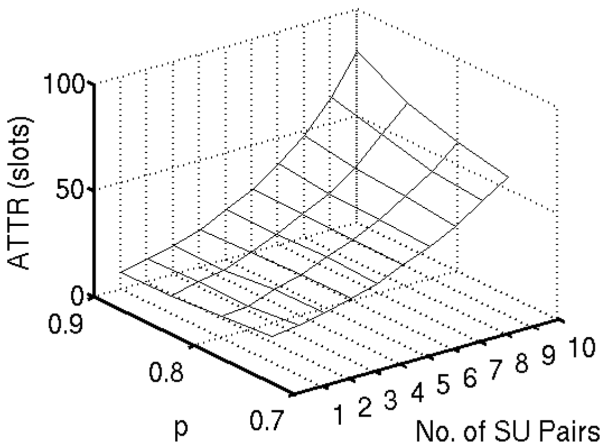
- SU’s use of channels result in slightly higher PUIR for MAQLR strategy.

Simulation Results (Cont'd)

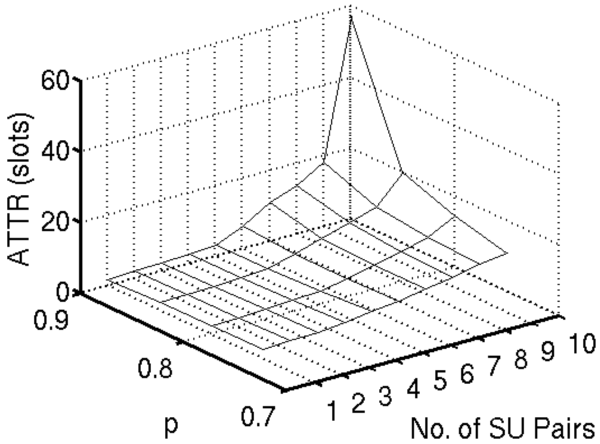
- **Average Time-to-Rendezvous (ATTR)**
  - **Average number of slots to complete RTS/CTS handshake**



EAMRCC-VS



NGQFH

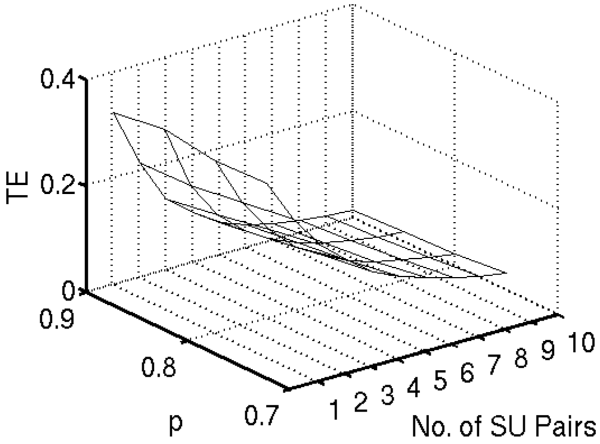


MAQLR

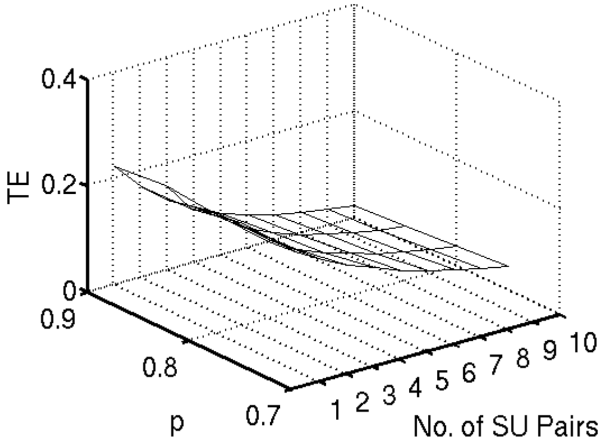
- MAQLR strategy has much lower ATTR mainly because of lower SUCR.

Simulation Results (Cont'd)

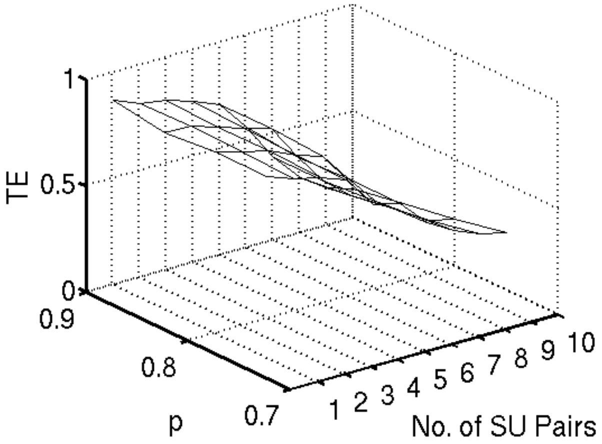
- **Throughput Efficiency (TE)**
  - **Ratio of actual throughput and maximum achievable throughput**
  - **Throughput is DATA packets exchanged per time slot.**



EAMRCC-VS



NGQFH



MAQLR

- MAQLR strategy has higher TE primarily because of lower SUCR.

Conclusion

- SUs enhance rendezvous performance with MAQLR strategy.
  - Actively learns which channels are best for rendezvous
  - Learns channels based on perceived PU activities and rendezvous successes/failures
  - Learns how to access channels effectively and efficiently
- Enhanced performance comes at the cost of higher PUIR.
- Future plans to improve strategy by lowering PUIR while still achieving desired (if not better) performance.

Questions