Detecting Sybil Nodes in Wireless Networks with Physical Layer Network Coding

Weichao Wang*, Di Pu**, and Alex Wyglinski**

*: SIS Dept., UNC Charlotte

**: ECE Dept., WPI

UNC CHARLOTTE

Motivation

- Network coding technique
  - improve network throughput, reduce congestion and enhance robustness
  - previous research focuses on the protection of NC and the detection of pollution attacks
- A different aspect: can network coding be used to detect malicious attacks?
  - Avoid the adoption of complex security schemes
  - Provide a new incentive for deployment of NC
  - Initial exploration in this paper: Sybil attacks in WN

Presentation organization

- Motivation
- Background
- Basic Idea
- Physical layer issues
- Network layer issues
- Analysis
- Related work
- Conclusions and future work

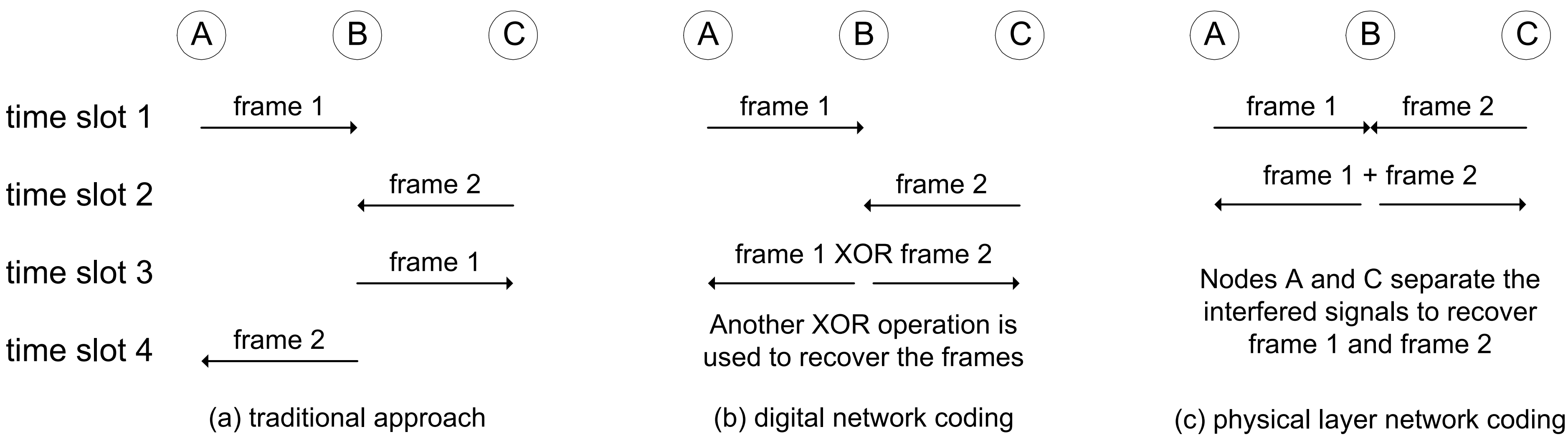
Background

- Sybil attacks in wireless networks
  - The same node presents multiple identities
  - is an example of stealth attack: difficult to detect through traditional methods
  - can threaten the safety of routing protocols and attack detection mechanisms
  - Previous Sybil detection schemes based on physical layer properties:
    - Depend on special hardware or inaccurate measurement

UNC CHARLOTTE

Background

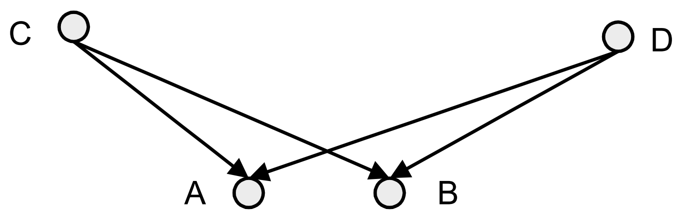
- PNC uses signal interference to achieve coding [MobiCom'06, SigComm'07]
- Not support random linear combination yet

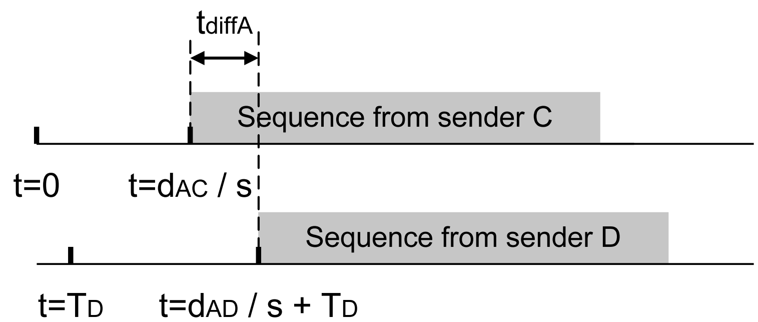


UNC CHARLOTTE

Basic idea

- The start point of signal interference is determined by the distances b/w the receivers and senders, and the sending time





- The difference b/w the arriving time at the receivers:

$$t_{diffA} = T_D + (d_{AD} - d_{AC}) / s$$

$$t_{diffB} = T_D + (d_{BD} - d_{BC}) / s$$

UNC CHARLOTTE

Basic idea

- The difference b/w two t_{diff} can cancel out the impacts of the sending time T_D

$$
\begin{aligned}
&\| t_{diffB} - t_{diffA} \| = \| (d_{BD} - d_{AD}) + (d_{AC} - d_{BC}) \| / s \\
&\leq (\| d_{BD} - d_{AD} \| + \| d_{AC} - d_{BC} \|) / s \;\leq\; 2 \times d_{AB} / s
\end{aligned}
$$

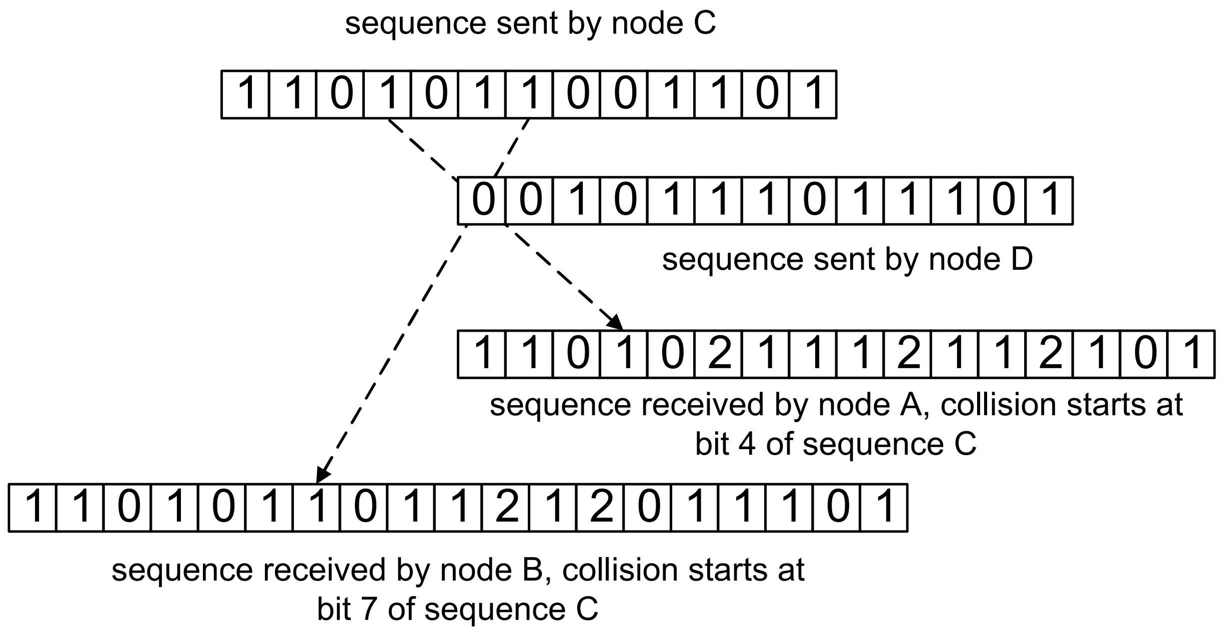
- The difference b/w t_{diffA} and t_{diffB} is restricted by the distance b/w A and B.
- If A and B are two physical nodes, they will demonstrate different time differences under different sender pairs
- If A and B are linked to the same physical node, they will always receive the same interference sequences

Basic idea

- Therefore, we can detect the Sybil nodes by examining the interference sequences at the nodes
- A mechanism is needed to verify the time difference
  - Cannot directly ask the nodes for their time difference: the Sybil nodes will lie to avoid detection
  - If $|| t_{diffA} – t_{diffB} ||$ is large enough, the two nodes can combine their received signals to recover the two sequences
  - The Sybil nodes will always get the same interference results and cannot separate the sequences

UNC CHARLOTTE

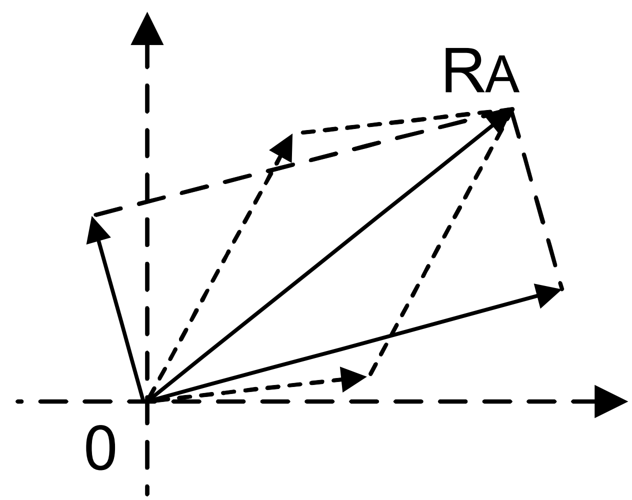
Basic idea



- Advantages: no synchronized clocks, no special hardware, distributed algorithm
- To turn the approach into a practical solution, efforts in both physical and network layers are needed

Physical layer issues

- Our approach is not bound to any signal modulation techniques; below MSK is assumed
  - Represent the data bits by varying the phase difference b/w consecutive signals
    - π/2 = bit "1", -π/2 = bit "0"
  - The receiver will get the vector sum of the two colliding signals



UNC CHARLOTTE

Physical layer issues

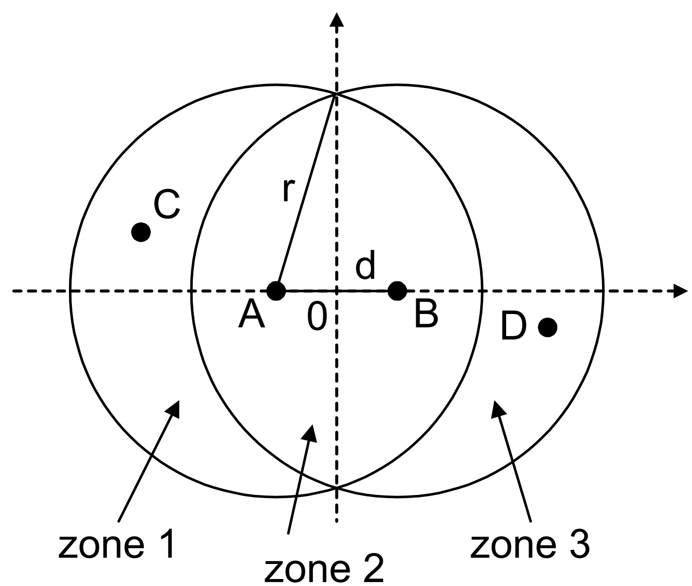
- Procedure to separate the colliding signals
  - Estimate the magnitudes of the two vectors [Katti et al. Sigcomm'07]
  - Use prior knowledge about one sequence or combine two different signal interference results to recover the data sequences
- Detect the start of signals and collisions
  - Use the incoming energy level changes to detect the first sequence
  - Measure the variance in the energy level of the incoming signals to detect collision

Network layer issues

- Network assumptions
  - Unit disk graph model for neighbor detection
  - Wireless nodes can adjust the transmission power
  - Share a secure, lightweight pseudo random bit generator
  - Omni-directional antenna
- The Sybil nodes
  - Have access to all knowledge bound to the identities under their control
  - Cannot compromise encryption keys or reverse a hash function

Network layer issues

- Selection of senders
  - Choose senders from the union of the neighbors of A and B: a pool much larger than the shared neighbors
  - The senders adjust the transmission power so that both receivers will get the signals



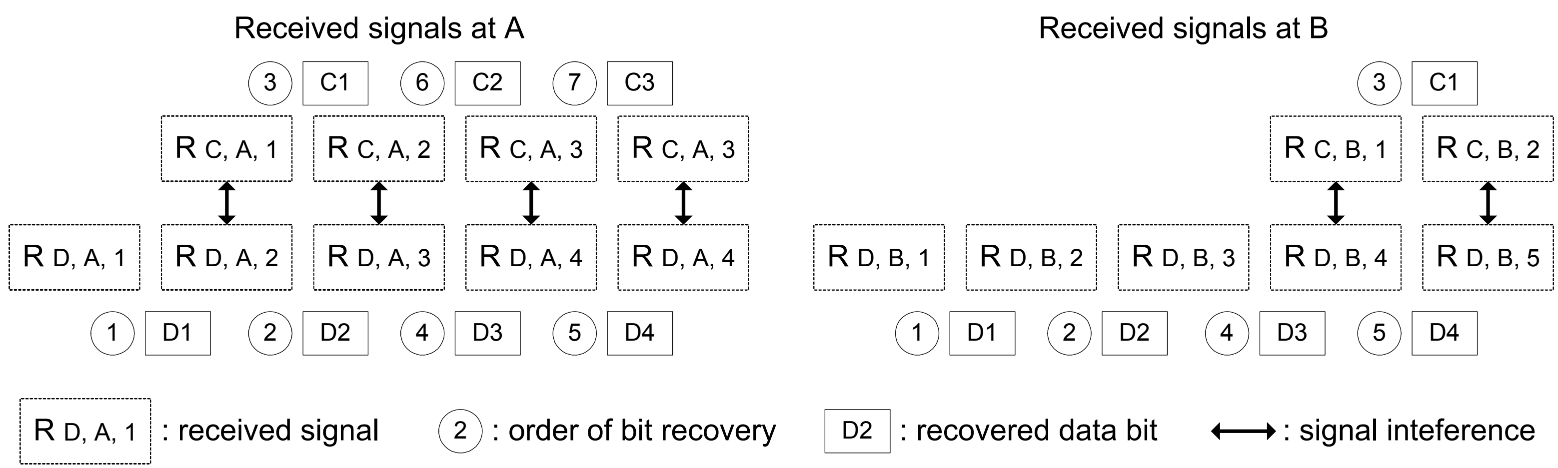
UNC CHARLOTTE

Network layer issues

- Generation of sending sequences
  - The sequences should satisfy two conditions:
    - Kept as a secret before they are sending out
    - Committed sequences and cannot be changed by the (malicious) senders
  - Sequence generation procedure
    - The senders select their seeds for the PRBG
    - The hash results of the seeds are broadcasted as the commitment of the sequences

Network layer issues

- Data recovery procedure
  - Under MSK modulation the receiver needs two signals to reconstruct one bit
  - Our analysis shows that when $\| t_{diffA} - t_{diffB} \| \geq 2$ signals, the two receivers can combine the interference signals to rebuild the sequences

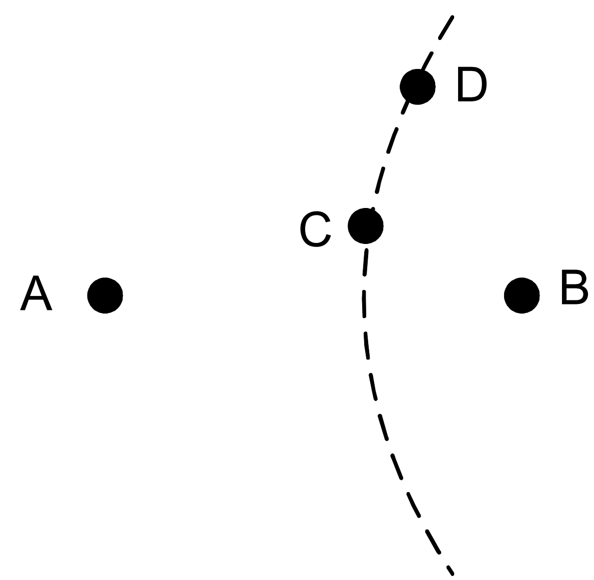


- Data recovery procedure
  - The receivers will broadcast the decoding results; the senders will broadcast the seeds
  - all nodes can verify the recovery results

UNC CHARLOTTE

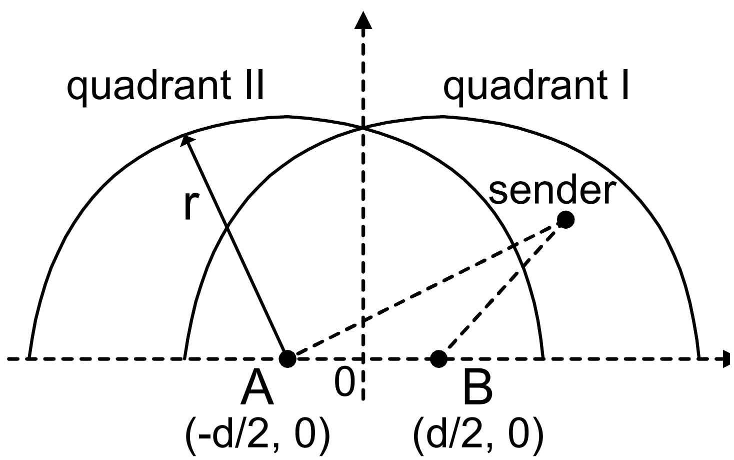
Analysis

- Handling false positive alarms
  - Even if the receivers are two different physical nodes, there is still a chance that they cannot reconstruct the packets
  - Example: two senders C and D are on the same hyperbola with the foci points A and B



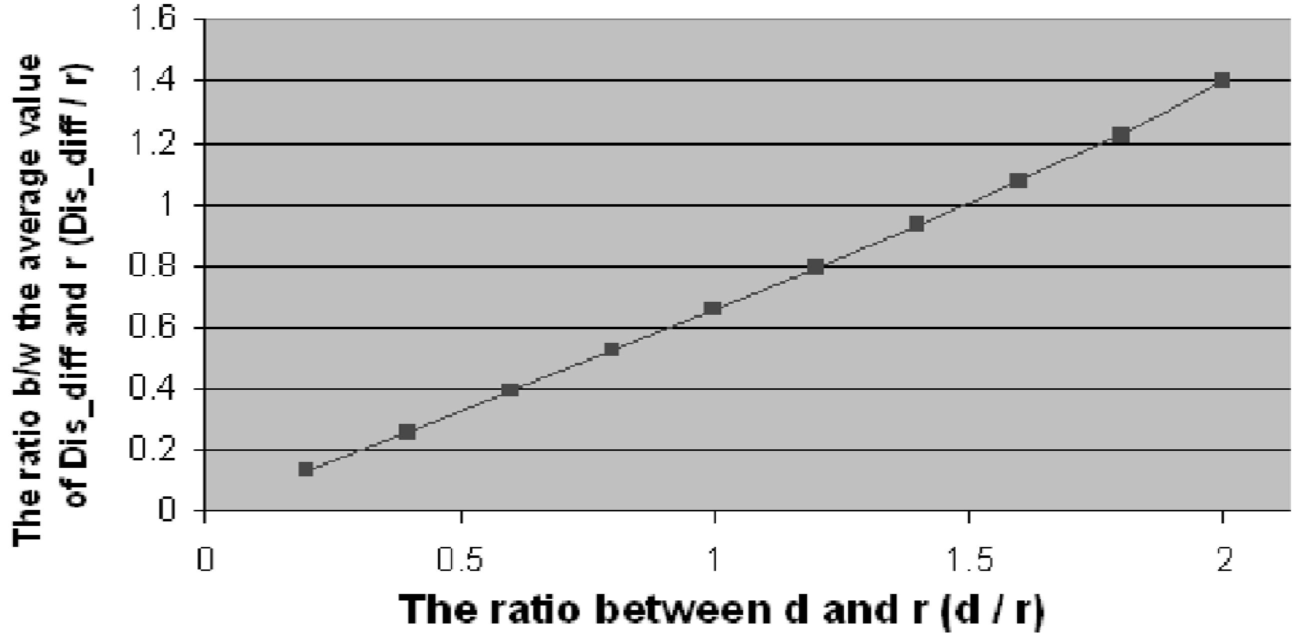
Analysis

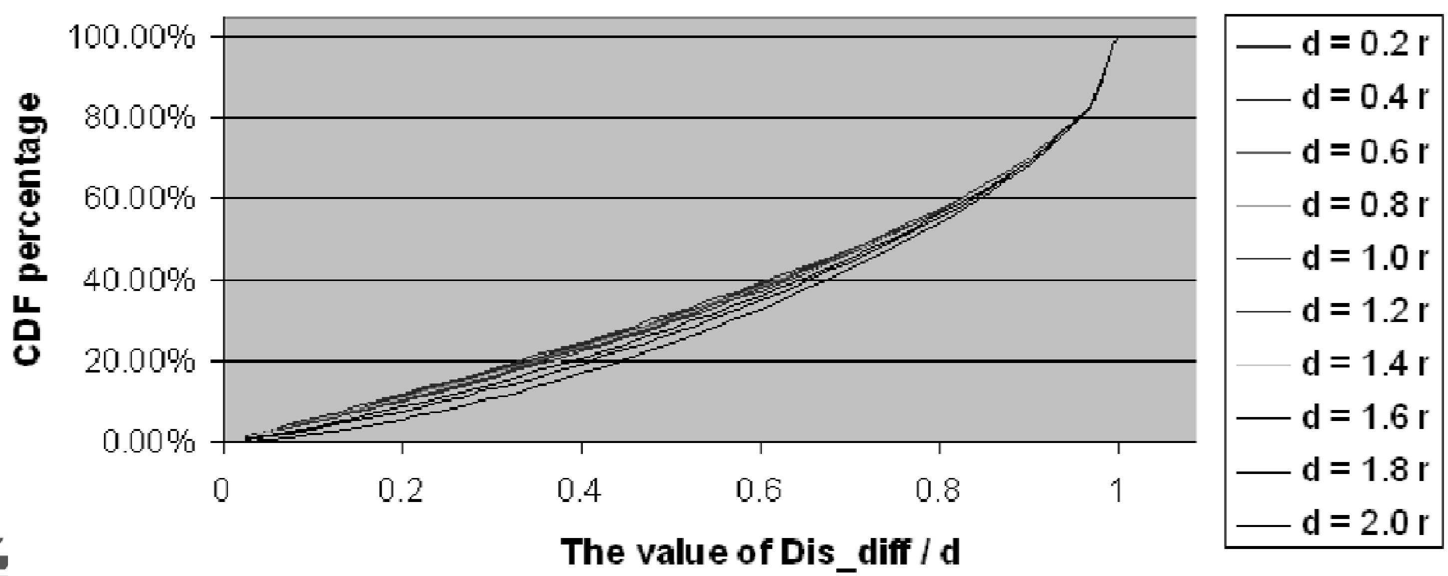
- Handling false positive alarms
  - An intuitive approach: multiple rounds of detection
  - We need a quantitative analysis



$$E[Dis_{diff}] = \frac{\int_{x=0}^{\frac{1}{2}d+r} \int_{y=0}^{\sqrt{r^2-(x-\frac{1}{2}d)^2}} Dis_{diff}\, dx\, dy}{area\ in\ Quadrant\ I}$$

$$= \frac{\int_{x=0}^{\frac{1}{2}d+r} \int_{y=0}^{\sqrt{r^2-(x-\frac{1}{2}d)^2}} Dis_{diff}\, dx\, dy}{\frac{1}{4} \cdot (2\pi r^2 - 2r^2 arccos(\frac{d}{2r}) + d\sqrt{r^2 - (\frac{d}{2})^2})}$$





UNC CHARLOTTE

Analysis

- Observations from the figures
  - The average value of Dis_{diff} has a nearly-constant ratio to d
  - From the CDF figure, the Dis_{diff} has a very low probability to have a small value
  - An empirical example
    - r=250m, d in [0, 2r], then $P[Dis_{diff} \leq 3m] \approx 0.01$
    - For one round of detection, when the senders are chosen from different sides of the Y-axis, $P[|| t_{diffA} - t_{diffB}|| \leq 3m / c] \leq 0.01\%$
    - Multiple rounds of detection will lead to a very low false positive detection rate

UNC CHARLOTTE

Analysis

- Why depend on PNC instead of system clocks to measure the time difference
  – The clock drift of wireless nodes is at micro-second level
  – The software defined-radio can easily use a much higher frequency
  – We will have a much higher Sybil detection sensitivity

UNC CHARLOTTE

Analysis

- Safety of the approach
  - When the selected senders are malicious
    - It is not easy for malicious senders to frame good receivers since they have committed to the sequences
    - If they are attached to the same physical node, all other nodes will receive the same interference results
    - They can disclose their sequences to Sybil nodes: multiple rounds of detection are needed
  - Frequency adjustment enabled by SDR
    - Control the Sybil detection accuracy
    - Avoid the jamming attacks

Related work

- Sybil detection
  - Identity based approaches
  - Location based approaches
  - Signal-print based approaches: measure RSSI at multiple positions [WiSe'06] or use radio signal transient shape [IPSN'09]
- Physical layer network coding
  - With synchronization at the senders [MobiCom'06]
  - Analog network coding [sigcomm'07]

Conclusions

- Exploring the security capabilities of Physical Layer Network Coding
- Using Sybil attack detection as a concrete example
- Advantages:
  - Avoid the dependence on special hardware
  - Take advantage of bandwidth efficiency improvement mechanisms
- Other potential applications
  - Localization [GlobeCom'10]
  - Other attacks on topology and identity

Limitations and future work

- What about attackers with multiple antennas or directional antennas
- What about collaborative attackers
- Implementation on SDR
- Thanks. Questions?