贝叶斯安全博弈

吴慧慈

北京邮电大学

E-mail：dailywu@bupt.edu.cn

引言

- In many security scenarios, the defenders and malicious attackers have limited information about each other. One of the main restrictions on available information in the security domain is the defenders' limited observation capabilities.
- Even if the defender has an accurate estimate of attacker preferences, the limitations on detection capabilities have to be taken into account as a factor on defensive decisions.
- the attackers can exploit their knowledge of imperfect detection when choosing their targets.
- All these considerations on limited observation and detection can be formalized within the framework of Bayesian security games.

引言

- Bayesian games model lack of information about the properties of players in a non-cooperative game using a probabilistic approach.
- In a Bayesian game, the players are usually assumed to be one of many specific types.
- A special *nature* player is introduced to the game which assigns a predetermined probability distribution to each player and type combination, which constitutes its fixed strategy.
- Subsequently, the original players compute their own strategies by taking into account each possible player-type combination weighted by the predetermined probability distribution.

Bayesian Intrusion Detection Game

- The defense systems often include a (virtual) sensor network in order to collect information and detect malicious attacks, which can be represented by the nature player in the Bayesian security game model.
- A virtual sensor network is defined as a collection of autonomous hardware and/or software agents that monitors the system and collects data, e.g. for detection purposes.
- The sensors report possible intrusions or anomalies occurring in a subsystem using common techniques such as signature comparison, pattern detection, or statistical analysis.

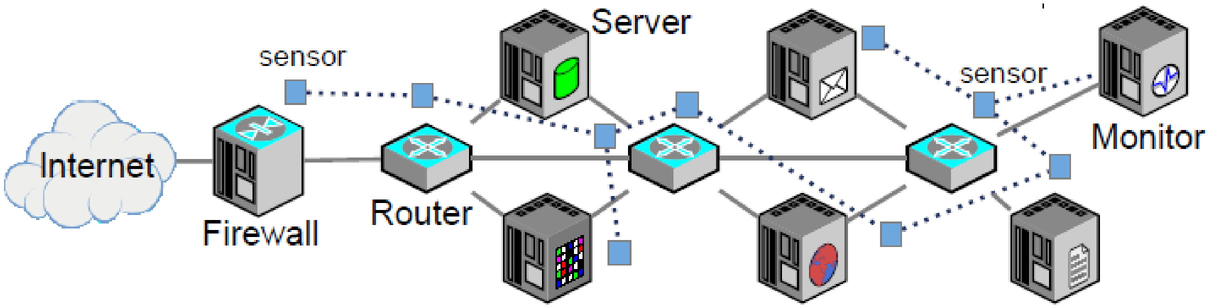
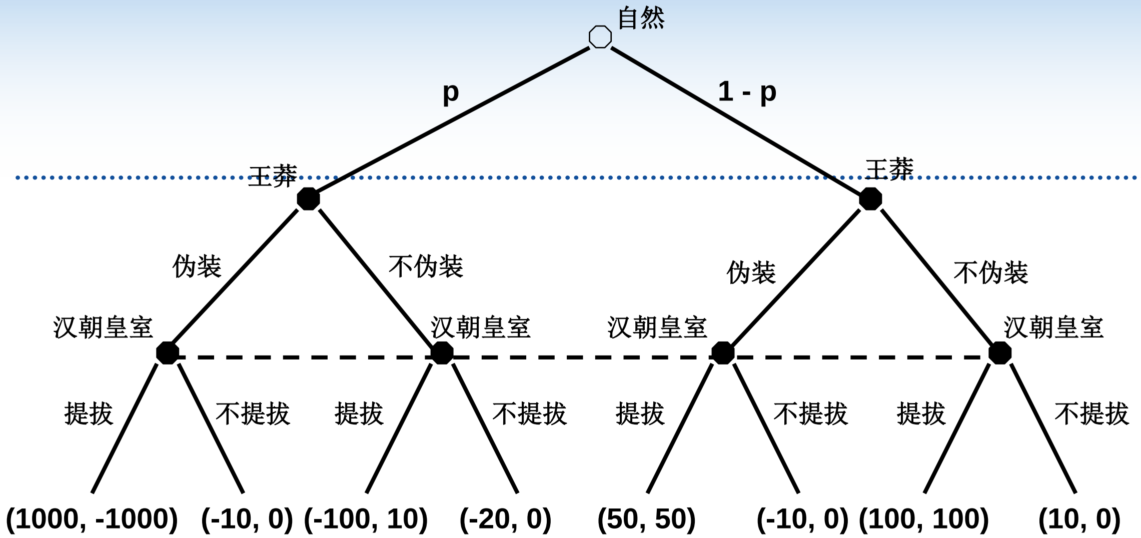


Figure 5.1 An intrusion detection system with a network of virtual sensors.



自然

p

1 - p

王莽

王莽

伪装

不伪装

伪装

不伪装

汉朝皇室

汉朝皇室

汉朝皇室

汉朝皇室

提拔

不提拔

提拔

不提拔

提拔

不提拔

提拔

不提拔

(1000, -1000)

(-10, 0)

(-100, 10)

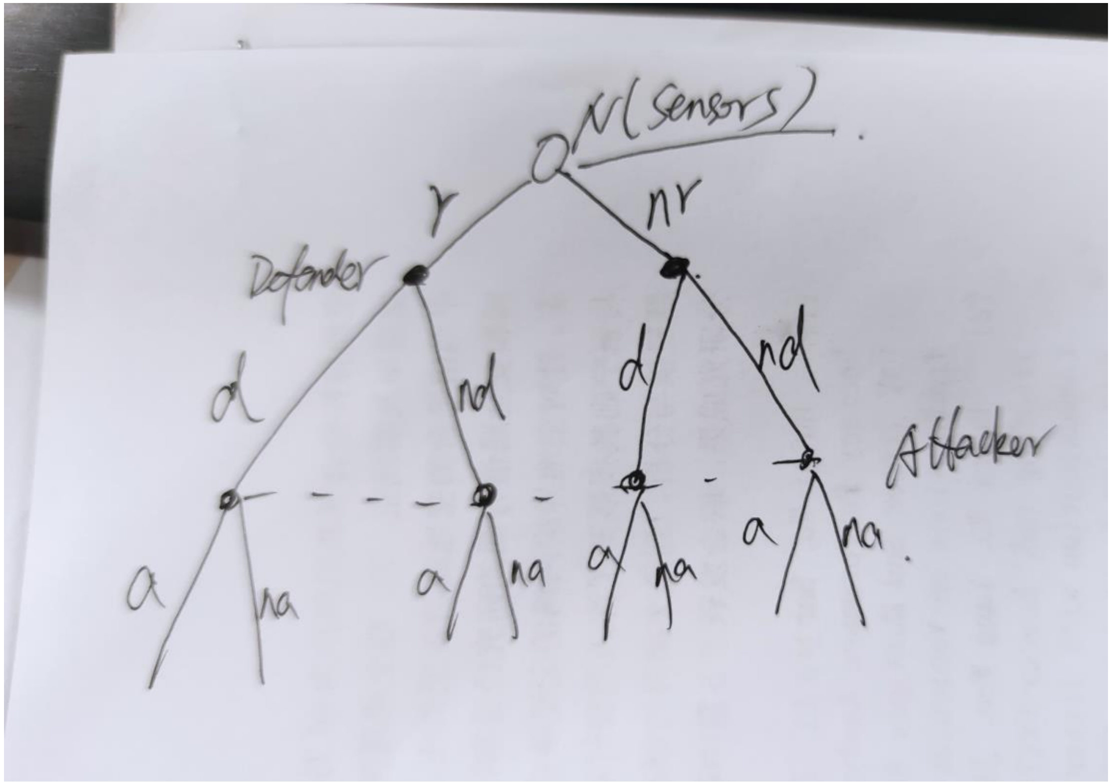
(-20, 0)

(50, 50)

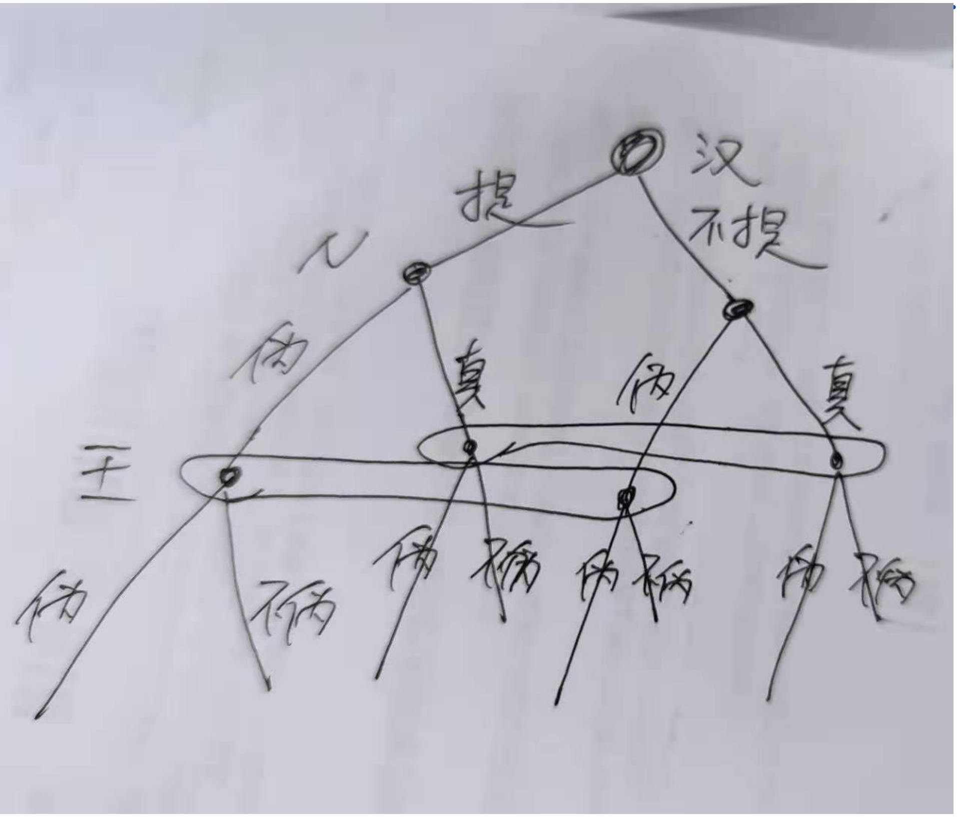
(-10, 0)

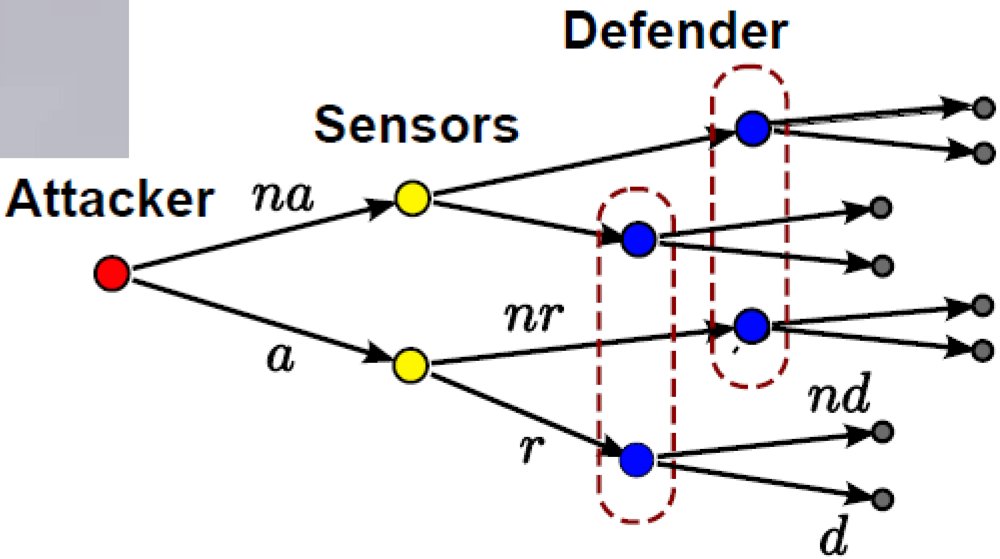
(100, 100)

(10, 0)



Bayesian Intrusion Detection Game





Bayesian Intrusion Detection Game

- The augmented set of attacks and anomalies over a target system T is $$\mathcal{A}^A := \{a_1, a_2, \ldots, a_{N_A}\} \cup \{na\},$$

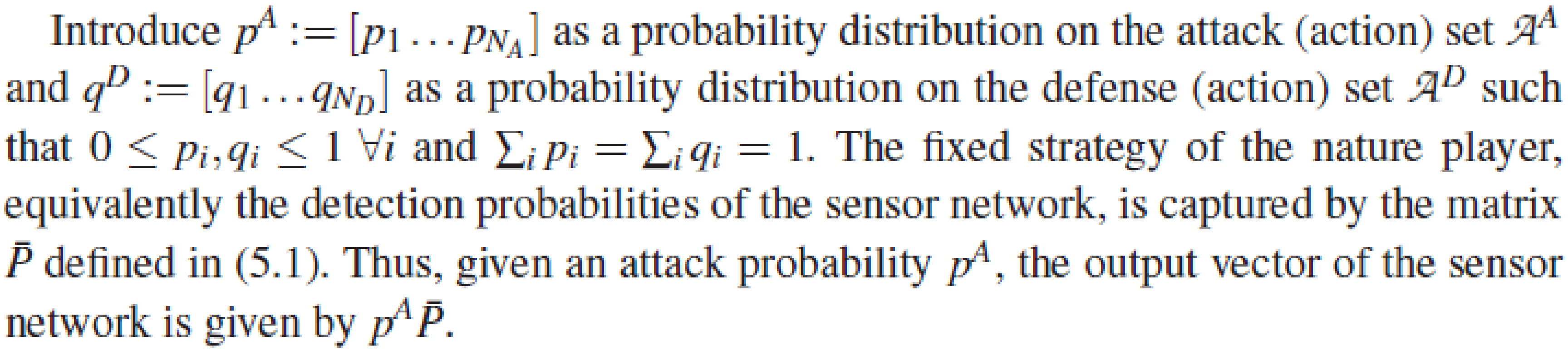
- Define linear mapping $\overline{P}: A^A \rightarrow A^A$ as the relationship between the actual attacks and the output of the sensor network S.

$$\bar{P} := [\bar{P}_{ij}]_{N_A \times N_A}, \text{ where } 0 \leq \bar{P}_{ij} \leq 1, \ \forall i, j \in \{1, \ldots, N_A\},$$

- The entry of the matrix $\overline{P}_{i,j}$ denotes the probability of attack i being reported as attack j. If $i \neq j$, then the sensor network confuses one attack for another. Such misreporting is quite beneficial for the attacker.

Bayesian Intrusion Detection Game

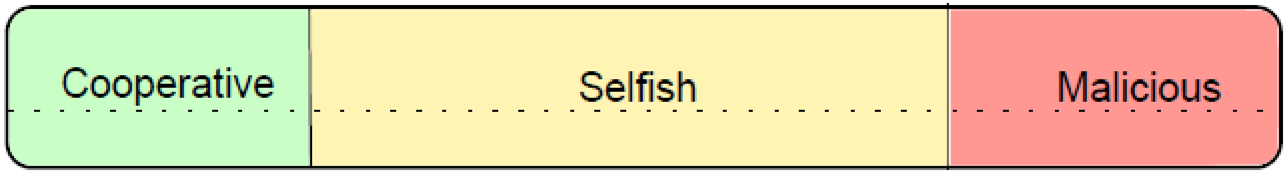
- In the case of j = na, $\overline{P}_{i,j}$ is the probability of missing to report an existing attack. Similarly, if i = na and $j \neq na$, then $\overline{P}_{i,j}$ is the probability of false alarm for attack j.
- Matrix $\overline{\mathbf{P}}$ describes the fixed strategy of the nature player (virtual sensor network).



Introduce $p^A := [p_1 \dots p_{N_A}]$ as a probability distribution on the attack (action) set $\mathcal{A}^A$ and $q^D := [q_1 \dots q_{N_D}]$ as a probability distribution on the defense (action) set $\mathcal{A}^D$ such that $0 \le p_i, q_i \le 1$ $\forall i$ and $\sum_i p_i = \sum_i q_i = 1$. The fixed strategy of the nature player, equivalently the detection probabilities of the sensor network, is captured by the matrix $\bar{P}$ defined in (5.1). Thus, given an attack probability p^A, the output vector of the sensor network is given by $p^A \bar{P}$.

Random Access Security Games

- Behavior type of different players starts with cooperation on the one end and ends with malicious behavior on the other



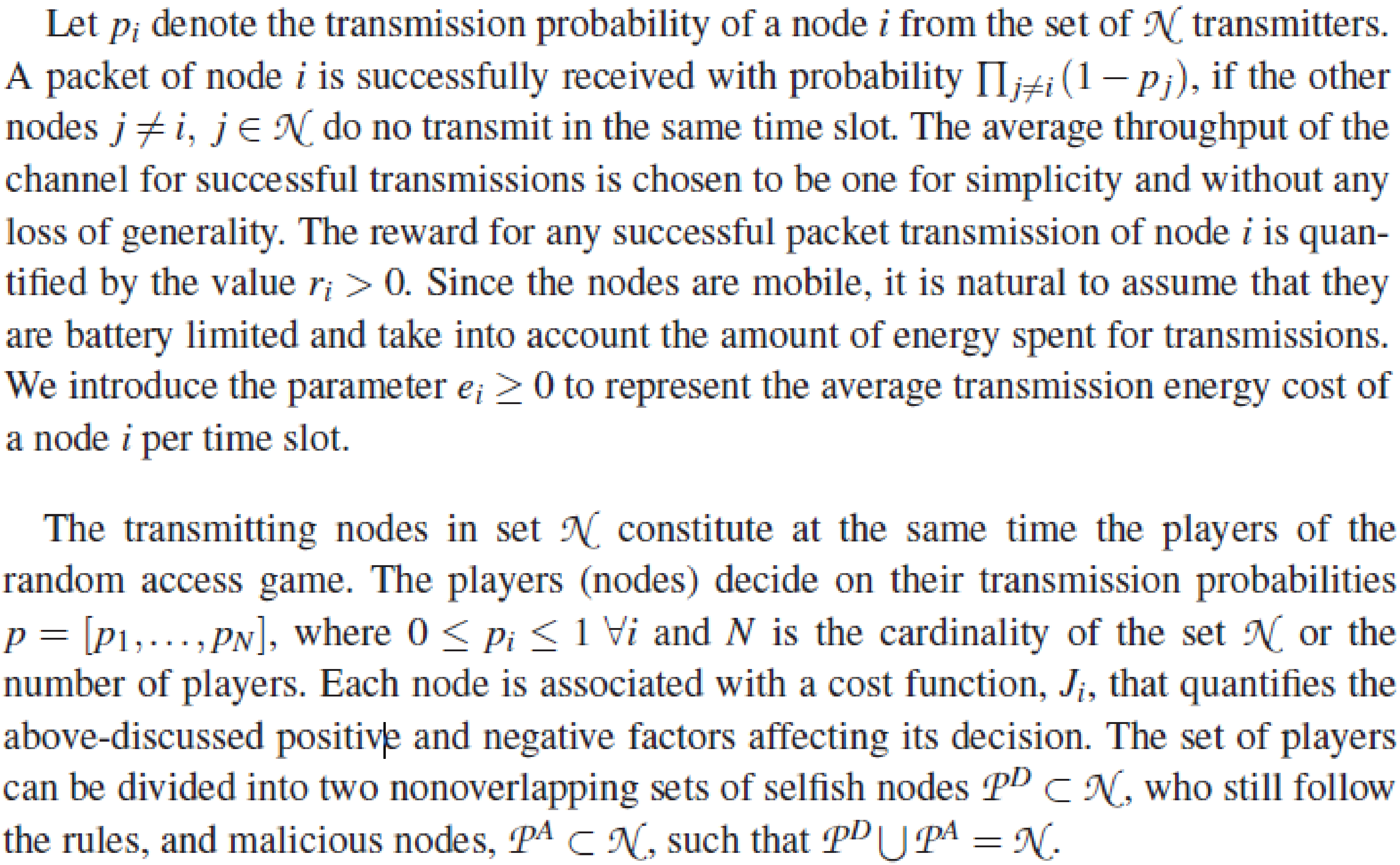
- The players of wireless security (jamming) games do not exactly know the type of other players, i.e. whether they are malicious or selfish, but have their own estimates. The Bayesian security games provide a framework for analyzing these more realistic scenarios.

Random Access Security Games

- The random access game models a wireless network with multiple nodes transmitting packets to a common receiver over a shared random access channel.
- Each transmitter is assumed to have saturated queues with uninterrupted availability of packets at any time slot and infinite buffer capacity for simplicity.
- The random access channel is a slotted system, in which each packet transmission takes one time slot.
- Any simultaneous transmission or collusion results in packet loss, which corresponds to a *classical collision channel*.

Random Access Security Games

❖ Perfect information

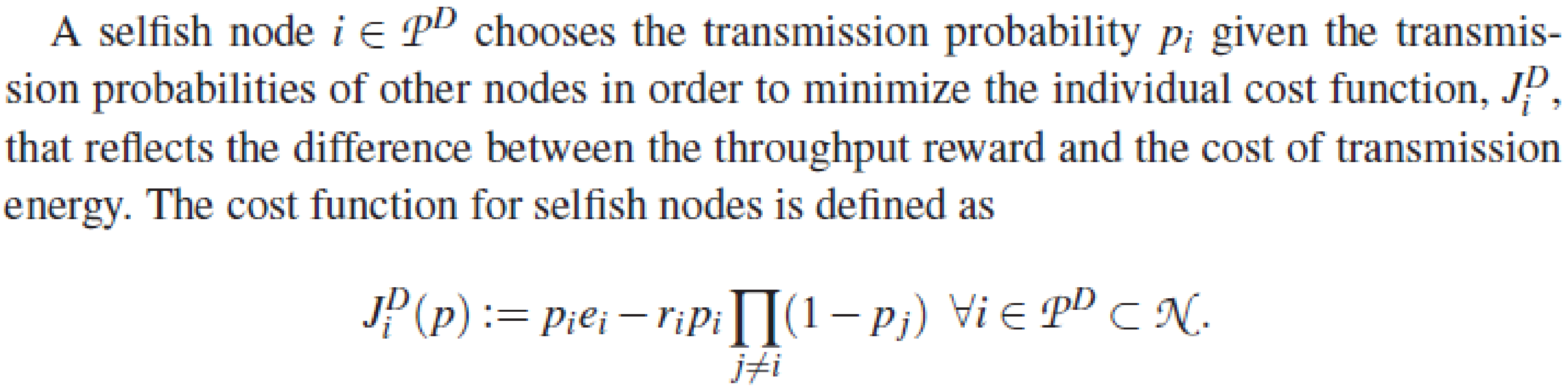


Let p_i denote the transmission probability of a node i from the set of $\mathcal{N}$ transmitters. A packet of node i is successfully received with probability $\prod_{j \neq i}(1 - p_j)$, if the other nodes $j \neq i$, $j \in \mathcal{N}$ do no transmit in the same time slot. The average throughput of the channel for successful transmissions is chosen to be one for simplicity and without any loss of generality. The reward for any successful packet transmission of node i is quantified by the value $r_i > 0$. Since the nodes are mobile, it is natural to assume that they are battery limited and take into account the amount of energy spent for transmissions. We introduce the parameter $e_i \geq 0$ to represent the average transmission energy cost of a node i per time slot.

The transmitting nodes in set $\mathcal{N}$ constitute at the same time the players of the random access game. The players (nodes) decide on their transmission probabilities $p = [p_1, \ldots, p_N]$, where $0 \leq p_i \leq 1$ $\forall i$ and N is the cardinality of the set $\mathcal{N}$ or the number of players. Each node is associated with a cost function, J_i, that quantifies the above-discussed positive and negative factors affecting its decision. The set of players can be divided into two nonoverlapping sets of selfish nodes $\mathcal{P}^D \subset \mathcal{N}$, who still follow the rules, and malicious nodes, $\mathcal{P}^A \subset \mathcal{N}$, such that $\mathcal{P}^D \bigcup \mathcal{P}^A = \mathcal{N}$.

Random Access Security Games

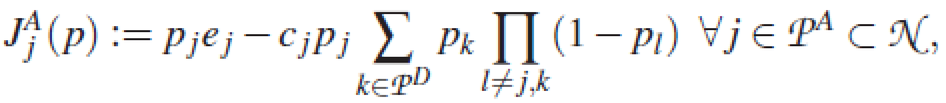
❖ Perfect information



A selfish node $i \in \mathcal{P}^D$ chooses the transmission probability p_i given the transmission probabilities of other nodes in order to minimize the individual cost function, J_i^D, that reflects the difference between the throughput reward and the cost of transmission energy. The cost function for selfish nodes is defined as

$$J_i^D(p) := p_i e_i - r_i p_i \prod_{j \neq i} (1 - p_j) \ \ \forall i \in \mathcal{P}^D \subset \mathcal{N}.$$

- The malicious nodes, unlike selfish ones, are motivated by the “reward” for disrupting or jamming the transmissions of other nodes rather than their own throughput. Hence, the cost function for a malicious node j is defined as



$$J_j^A(p) := p_j e_j - c_j p_j \sum_{k \in \mathcal{P}^D} p_k \prod_{l \neq j,k} (1 - p_l) \ \ \forall j \in \mathcal{P}^A \subset \mathcal{N},$$

Random Access Security Games

❖ Two selfish players

- When both players are selfish

$$J_1(p) := p_1 e_1 - r_1 p_1 (1 - p_2),$$

Theorem 3.2. *The random access game with two selfish players admits the following Nash equilibrium strategies (p_1^*, p_2^*) under the respective conditions.*

(a) If $0 < e_1 \leq r_1$ and $0 < e_2 \leq r_2$, there exist three Nash equilibria:

$$p_1^* = 1,\ p_2^* = 0;\ p_1^* = 0,\ p_2^* = 1;\ p_1^* = 1 - \frac{e_2}{r_2},\ p_2^* = 1 - \frac{e_1}{r_1}.$$

(b) If $e_2 = 0$ and $e_1 < r_1$, there exist a continuum of NE:

$$p_1^* = 1,\ p_2^* \in [1 - \frac{e_1}{r_1}, 1],$$

and if $e_1 = 0$ and $e_2 < r_2$, there exist likewise a continuum of NE:

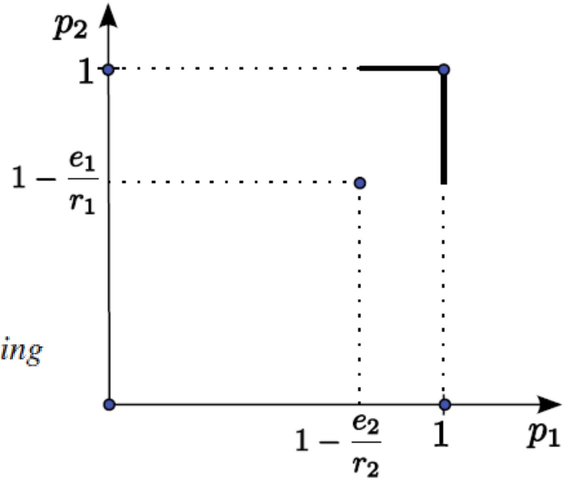
$$p_1^* \in [1 - \frac{e_2}{r_2}, 1],\ p_2^* = 1.$$

(c) If $e_1 > r_1$ or $e_2 > r_2$ or both, the NE is unique:

(i) $p_1^* = p_2^* = 0$, if $e_1 > r_1$ and $e_2 > r_2$.

(ii) $p_1^* = 0$, $p_2^* = 1$, if $e_1 > r_1$ and $e_2 < r_2$.

(iii) $p_1^* = 1$, $p_2^* = 0$, if $e_1 < r_1$ and $e_2 > r_2$.

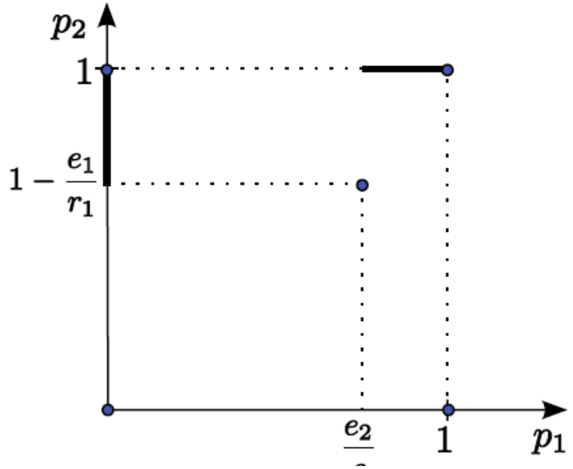


quare $0 \leq p_1 \leq 1$, $0 \leq$
ner NE. Since both J_1
$p_1^*, p_2^*)$ has to have the
dependent of p_2, which
ided that $0 < e_1 \leq r_1$,

in this solution which
) and $e_2 = 0$, which are
ave to be at the corners
theorem, which follow
□

Random Access Security Games

❖ One selfish player and one malicious playe



$$J_1(p) := p_1 e_1 - r_1 p_1 (1 - p_2),$$
$$J_2(p) := p_2 e_2 - c p_2 p_1.$$

Theorem 3.3. *The random access game with one selfish player 1 and a malicious player 2 admits the following Nash equilibrium strategies (p_1^*, p_2^*) under the respective conditions.*

(i)

$$p_1^* = \frac{e_2}{c},\ p_2^* = 1 - \frac{e_1}{r_1},\ \text{if } 0 < e_1 \le r_1 \text{ and } 0 < e_2 \le c.$$

(ii) $p_1^* = p_2^* = 0$, *if* $e_1 > r_1$.

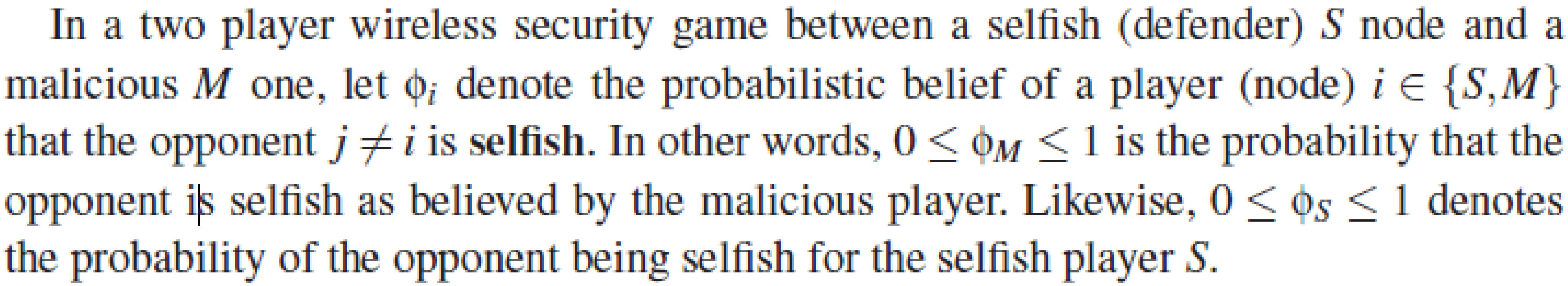
(iii) $p_1^* = 0$, $p_2^* \in [1 - \frac{e_1}{r_1}, 1]$, *if* $e_1 \le r_1$ *and* $e_2 = 0$.

(iv) $p_1^* = 1$, $p_2^* = 0$, *if* $e_1 < r_1$ *and* $e_2 > c$.

(v) $p_1^* \in [\frac{e_2}{c}, 1]$, $p_2^* = 1$, *if* $e_1 = 0$ *and* $e_2 \le c$.

Random Access Security Games

Perfect information



In a two player wireless security game between a selfish (defender) S node and a malicious M one, let ϕ_i denote the probabilistic belief of a player (node) $i \in \{S,M\}$ that the opponent $j \neq i$ is **selfish**. In other words, $0 \leq \phi_M \leq 1$ is the probability that the opponent is selfish as believed by the malicious player. Likewise, $0 \leq \phi_S \leq 1$ denotes the probability of the opponent being selfish for the selfish player S.

- the selfish node S receives unit throughput reward for successful transmission, i.e. r = 1. Then, the cost of the selfish node,

$$J_S(p) := p_S e_S - p_S(1-p_j),\ j \in \{S,M\}.$$

- Expected costs of players under the given probabilistic beliefs about the opponents

$$J_S(p_S, p_M) = p_S e_S - \phi_S p_S (1-p_S) - (1-\phi_S) p_S (1-p_M)$$

- Assume a malicious node M incurs a unit reward, if the opponent is selfish and successfully jammed at the given time slot, i.e. c = 1. The cost function of the malicious node,

$$J_M(p) := \begin{cases} p_M e_M - p_M p_S, & \text{if the opponent is selfish} \\ p_M e_M, & \text{if the opponent is malicious.} \end{cases}$$

- Expected costs $\quad J_M(p_S, p_M) := p_M e_M - \phi_M p_M p_S,$

Random Access Security Games

$J_S(p_S, p_M) = p_S e_S - \phi_S p_S (1 - p_S) - (1 - \phi_S) p_S (1 - p_M)$

- is minimized by transmitting with probability $p_S = 0$ if $e_S > \phi_S + (1 - \phi_S)(1 - p_M)$

- When $p_S = 0$, $J_M(p_S, p_M) = p_M e_M$ is minimized when $p_M = 0$.
- When $p_M = 0$, $p_S = 0$ is still the best response for S if $e_S > 1$
  - NE $p_S = 0, p_M = 0$ if $e_S > 1$

- is minimized by transmitting with probability $p_S = 1$ if $e_S < \phi_S + (1 - \phi_S)(1 - p_M)$

- When $p_S = 1$, $J_M(p_S, p_M) = p_M e_M - p_M \phi_M$ is minimized when $p_M = 0$ if $e_M > \phi_M$
- When $p_M = 0$, $p_S = 1$ is still the best response for S if $e_S < 1$
  - NE $p_S = 1, p_M = 0$ if $0 < e_S < 1$

- When $p_S = 1$, $J_M(p_S, p_M) = p_M e_M - p_M \phi_M$ is minimized when $p_M = 1$ if $e_M < \phi_M$
- When $p_M = 1$, $p_S = 1$ is still the best response for S if $e_S < \phi_S$
  - NE $p_S = 1, p_M = 1$ if $e_S < \phi_S$

Interference Limited Multiple Access Security Games

- Define $P_i \geq 0$ and $e_i \geq 0$ as the transmission power level and the corresponding energy cost (per unit power) of a node i, respectively.
- Multiple access security games within an interference-limited multi access scheme where the nodes choose their transmission power levels in order to maximize their signal-to-interference-plus-noise-ratio (SINR)

$$\gamma_i = \frac{h_i P_i}{\frac{1}{L}\sum_{j \neq i} h_j P_j + \sigma^2},$$

- Each selfish node independently chooses the power P_i for transmitting to a common receiver in order to minimize the individual expected cost J_i .

$$J_i(P) = e_i P_i - \gamma_i = e_i P_i - \frac{h_i P_i}{\frac{1}{L}\sum_{j \neq i} h_j P_j + \sigma^2},$$

Interference Limited Multiple Access Security Games

- Malicious nodes receive rewards for jamming others (in terms of decreasing their performance or SINR) rather than improving their own SINR levels.
- The cost function of a malicious node is defined as

$$J_i(P) = e_i P_i + \sum_{j \in \mathcal{P}^D} \gamma_j,$$

set of selfish nodes the malicious one targets

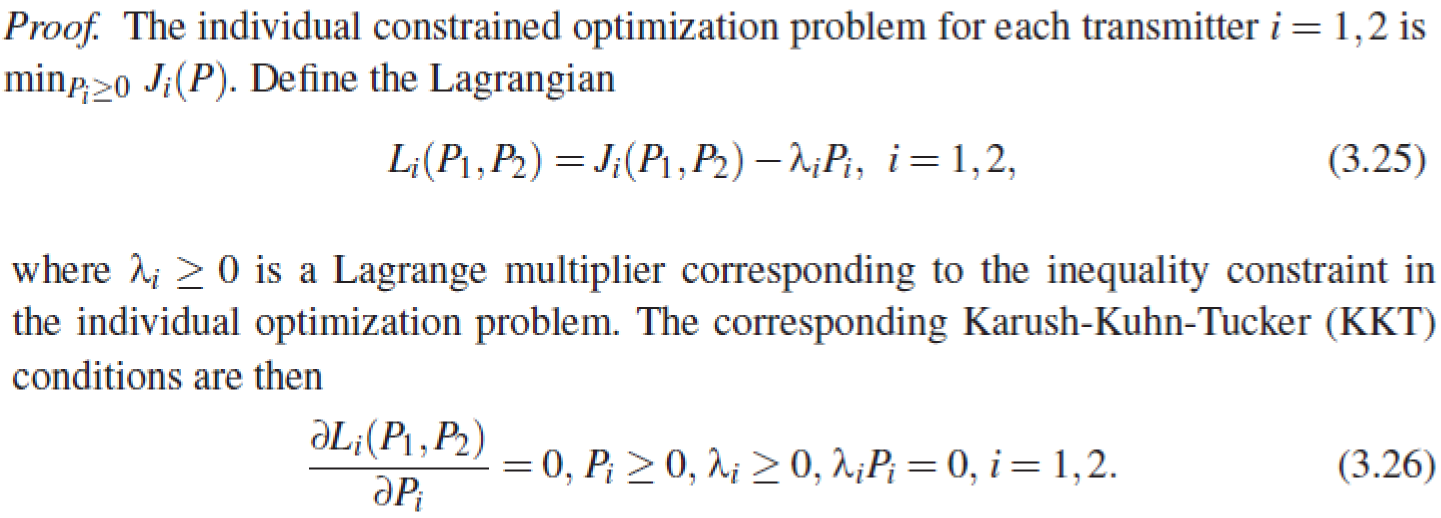
- the set of malicious nodes is denoted by P_A

Interference Limited Multiple Access Security Games

- Two selfish transmitters

$$\min_{P_1 \geq 0} J_1(P_1, P_2) = e_1 P_1 - \frac{h_1 P_1}{h_2 P_2 + \sigma^2},$$

$$\min_{P_2 \geq 0} J_2(P_1, P_2) = e_2 P_2 - \frac{h_2 P_2}{h_1 P_1 + \sigma^2}$$

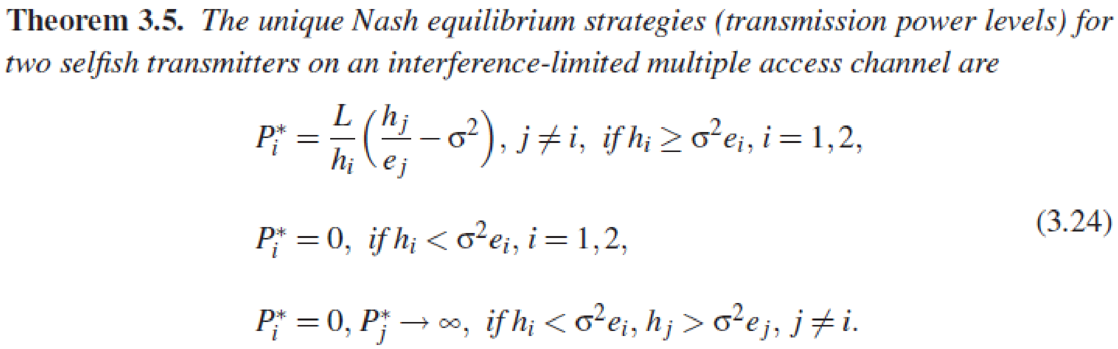


Proof. The individual constrained optimization problem for each transmitter $i = 1,2$ is $\min_{P_i \geq 0} J_i(P)$. Define the Lagrangian

$$L_i(P_1, P_2) = J_i(P_1, P_2) - \lambda_i P_i, \quad i = 1,2, \tag{3.25}$$

where $\lambda_i \geq 0$ is a Lagrange multiplier corresponding to the inequality constraint in the individual optimization problem. The corresponding Karush-Kuhn-Tucker (KKT) conditions are then

$$\frac{\partial L_i(P_1, P_2)}{\partial P_i} = 0,\ P_i \geq 0,\ \lambda_i \geq 0,\ \lambda_i P_i = 0,\ i = 1,2. \tag{3.26}$$



Theorem 3.5. *The unique Nash equilibrium strategies (transmission power levels) for two selfish transmitters on an interference-limited multiple access channel are*

$$\begin{aligned} &P_i^* = \frac{L}{h_i}\left(\frac{h_j}{e_j} - \sigma^2\right),\ j \neq i,\ \text{if}\ h_i \geq \sigma^2 e_i,\ i = 1,2, \\ &P_i^* = 0,\ \text{if}\ h_i < \sigma^2 e_i,\ i = 1,2, \\ &P_i^* = 0,\ P_j^* \to \infty,\ \text{if}\ h_i < \sigma^2 e_i,\ h_j > \sigma^2 e_j,\ j \neq i. \end{aligned} \tag{3.24}$$

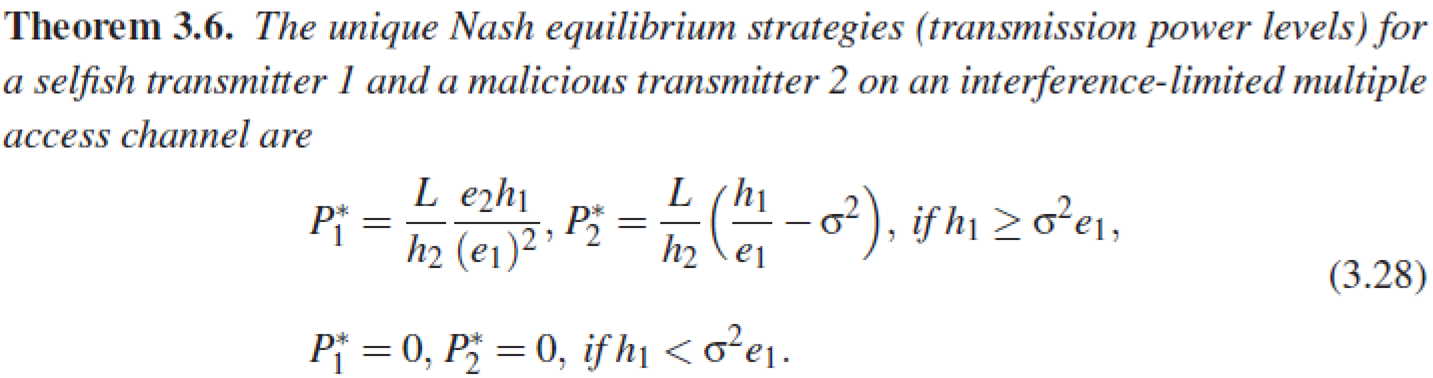
Interference Limited Multiple Access Security Games

❖ One malicious transmitter and one selfish transmitter

$$\min_{P_1 \geq 0} J_1\left(P_1, P_2\right) = e_1 P_1 - \frac{h_1 P_1}{h_2 P_2 + \sigma^2},$$

$$\min_{P_2 \geq 0} J_2\left(P_1, P_2\right) = e_2 P_2 + \frac{h_1 P_1}{h_2 P_2 + \sigma^2}$$

$\Downarrow$ KKT



Theorem 3.6. *The unique Nash equilibrium strategies (transmission power levels) for a selfish transmitter 1 and a malicious transmitter 2 on an interference-limited multiple access channel are*

$$P_1^* = \frac{L}{h_2} \frac{e_2 h_1}{(e_1)^2}, \; P_2^* = \frac{L}{h_2}\left(\frac{h_1}{e_1} - \sigma^2\right), \; \text{if } h_1 \geq \sigma^2 e_1, \tag{3.28}$$

$$P_1^* = 0, \; P_2^* = 0, \; \text{if } h_1 < \sigma^2 e_1.$$

Interference Limited Multiple Access Security Games

❖ Imperfect information

- The first player is again selfish and believes that the second one is also selfish with probability ϕ_1 . The second player can be indeed selfish or malicious, however regardless of his type he knows that the first player is selfish.
- Define P_{2S}, P_{2M} as the power levels and e_S, e_M as the energy costs of the second player for the selfish and malicious cases, respectively.

$$\min_{P_1 \geq 0} J_1(P_1, P_2) = e_S P_1 - \phi_1 \frac{h_1 P_1}{h_2 P_{2S} + \sigma^2} - (1-\phi_1) \frac{h_1 P_1}{h_2 P_{2M} + \sigma^2},$$

$$\min_{P_2 \geq 0} J_2(P_1, P_2) = e_S P_{2S} - \frac{h_2 P_{2S}}{h_1 P_1 + \sigma^2} \; or \; \min_{P_2 \geq 0} J_2(P_1, P_2) = e_M P_{2M} + \frac{h_1 P_1}{h_2 P_{2M} + \sigma^2}$$

$$P_1^* = \frac{L}{h_1}\left(\frac{h_2}{e_S} - \sigma^2\right), \qquad (5.20)$$

$$P_{2S}^* = \frac{L}{h_2}\left[\frac{h_1 \phi_1}{\left(e_S - \frac{(1-\phi_1)h_1}{\frac{h_2}{L}P_{2M} + \sigma^2}\right)} - \sigma^2\right]^+, \qquad (5.21)$$

$$P_{2M}^* = \left[\sqrt{\frac{L h_1 P_1^*}{h_2 e_M}} - \frac{L\sigma^2}{h_2}\right]^+, \qquad (5.22)$$

if $h_2 \geq \sigma^2 e_S$, *where* $[x]^+ = \max(x, 0)$.
Otherwise, that is if $h_2 < \sigma^2 e_S$, *then*

$$P_1^* = \left(\left[e_S - \frac{\phi_1 h_1}{\sigma^2}\right]^+\right)^{-2} \frac{e_M L h_1}{h_2} (1-\phi_1)^2, \qquad (5.23)$$

$P_{2S}^* = 0$ *and* P_{2M}^* *is given by (5.22) with* P_1^* *from (5.23).*