

Given

- a water distribution network  $W = (V, E)$  with
  - a set of  $N$  nodes (junctions, tanks, reservoirs)  $V = \{v_n | n \in \{1, \dots, N\}\}$
  - a set of edges (pipes, valves, pumps) connecting the nodes:  $E \subseteq V \times V$
- a number of timesteps  $T \in \mathbb{N}$
- pressure values for each node  $v_n$  given as realizations  $\mathbf{x}_n := (x_{n,1}, \dots, x_{n,T})$  of random variables  $\mathcal{X}_n \forall n \in \{1, \dots, N\}$  with underlying joint probability density function  $p(\mathcal{X})$
- A number  $S \leq N$  of nodes, which are equipped with sensors. The nodes can be partitioned into  $\{v_1, \dots, v_S\}$  which are equipped with sensors and  $\{v_{S+1}, \dots, v_N\}$  which have no sensor. In particular, this means that  $\mathcal{X}_{obs} := \{\mathcal{X}_1, \dots, \mathcal{X}_S\}$  are observed random variables, while  $\mathcal{X}_{lat} := \{\mathcal{X}_{S+1}, \dots, \mathcal{X}_N\}$  are latent.
- a prediction function  $f_{pred} : \mathbb{R}^{S \times K} \mapsto \{0, 1\}$  taking a matrix of pressure values  $\mathbf{X}_{obs}^{K+} := ((x_{1,T-K+1}, \dots, x_{1,T}), \dots, (x_{S,T-K+1}, \dots, x_{S,T}))$  for the  $K$  most recent timesteps as input and indicating whether the network is under attack (1) or not (0)

I will define the least sensitive point of the network as the node, for which the pressure values could deviate the most from their expected value while keeping the probability that the prediction function will issue an alarm due to this deviation below a fixed threshold of  $\frac{1}{2}$ . For this purpose, I will investigate the expected value of the outcome of the prediction function, based on the conditional probability distribution of all other nodes given the deviating node.

Let  $v_n$  with corresponding random variable  $\mathcal{X}_n$  be the node, for which deviating pressure values shall be examined and let  $\mathbf{x}_n^{K+} := (x_{n,T-K+1}, \dots, x_{n,T})$  be a vector containing the last  $K$  realizations of  $\mathcal{X}_n$ . Let  $\mathcal{X}_{-n} := \{\mathcal{X}_i | i \in \{1, \dots, N\}, i \neq n\}$  be the random variables of all other nodes and  $\mathbf{X}_{-n}^{K+}$  be a matrix containing the last  $K$  realizations of  $\mathcal{X}_{-n}$ . Then the maximum tolerable deviation for a node  $v_n$  can be defined as

$$\text{maxdev}(v_n) := \max_{\mathbf{x}_n^{K+} \in \mathbb{R}^K} \|\mathbf{x}_n^{K+} - E[\mathcal{X}_n]\| \quad (1)$$

$$\text{s.t. } E_{p(\mathcal{X}_{-n}|\mathcal{X}_n)}[f_{pred}(\mathcal{X}_{obs})] < \frac{1}{2} \quad (2)$$

where

$$E_{P(\mathcal{X}_{-n}|\mathcal{X}_n)}[f_{pred}(\mathcal{X}_{obs})] = \int P(\mathcal{X}_{-n} = \mathbf{X}_{-n}^{K+} | \mathcal{X}_n = \mathbf{x}_n^{K+}) f_{pred}(\mathbf{X}_{obs}^{K+}) d\mathbf{X}_{-n}^{K+} \quad (3)$$

The least sensitive node is now given as the node with highest maximum tolerable deviation

$$\text{lsp}(V) := \arg \max_{v_n \in V} \text{maxdev}(v_n) \quad (4)$$