Bayesian Decision Aggregation in Collaborative Intrusion Detection Networks

Carol Fung, Quanyan Zhu,
Raouf Boutaba, and Tamer Başar
David Cheriton School of Computer Science,
University of Waterloo
Department of Electrical and Computer Engineering
University of Illinois at Urbana Champaign

Motivation

- Cyber intrusions are more sophisticated and harder to detect
  - Malware, botnet, DDoS
- Intrusion Detection System (IDS)
  - Compare computer activity/traffic with known intrusion patterns
  - Host-based and network-based
  - Can not cover all types of intrusions
  - Easily compromised by unknown or new threats
- An Collaborative Intrusion Detection Network (CIDN) allows IDSes to share knowledge and experience with others
  - Cover more intrusion types
  - Achieve higher detection accuracy

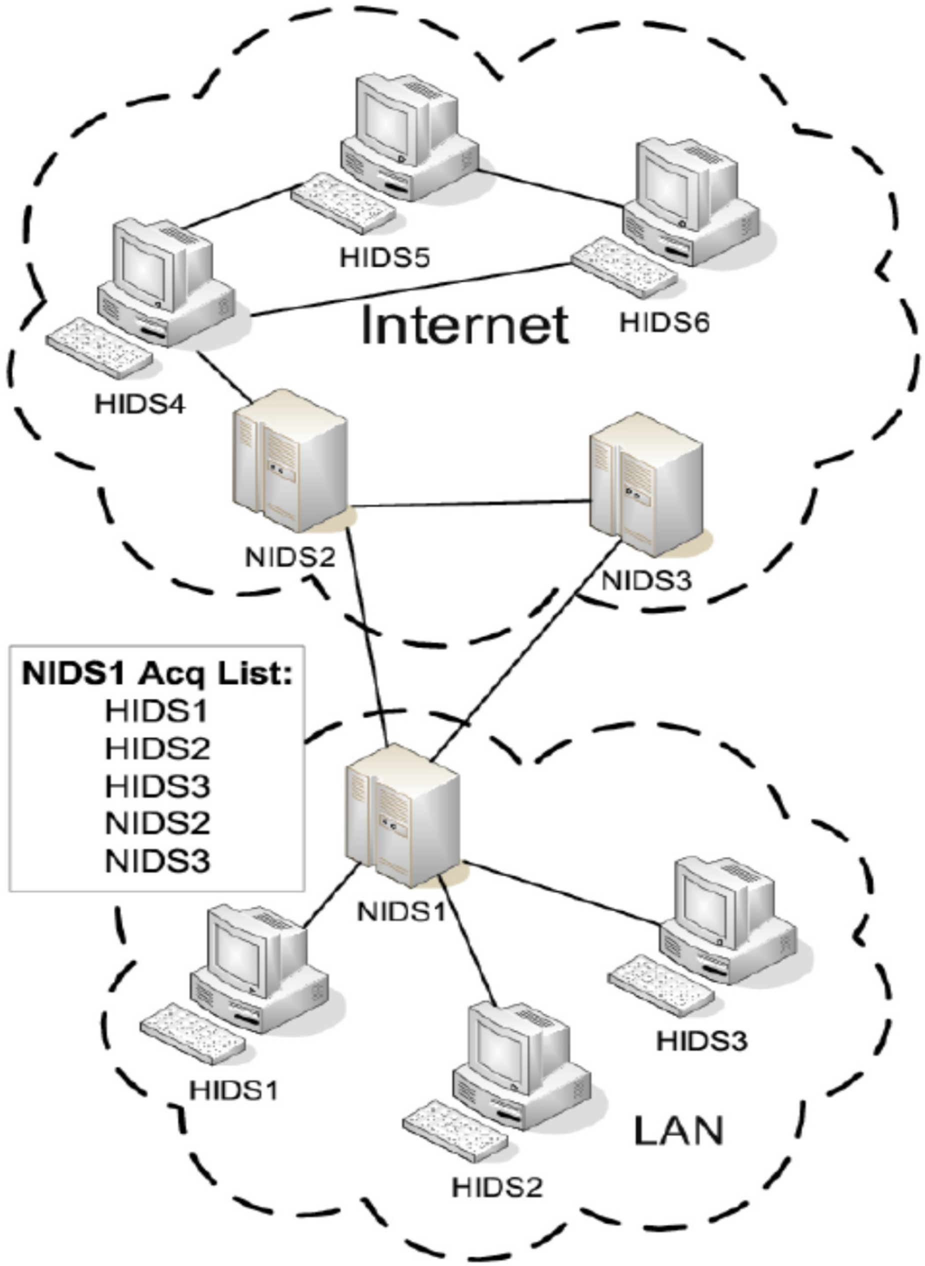


Figure 1. CIDN Topology

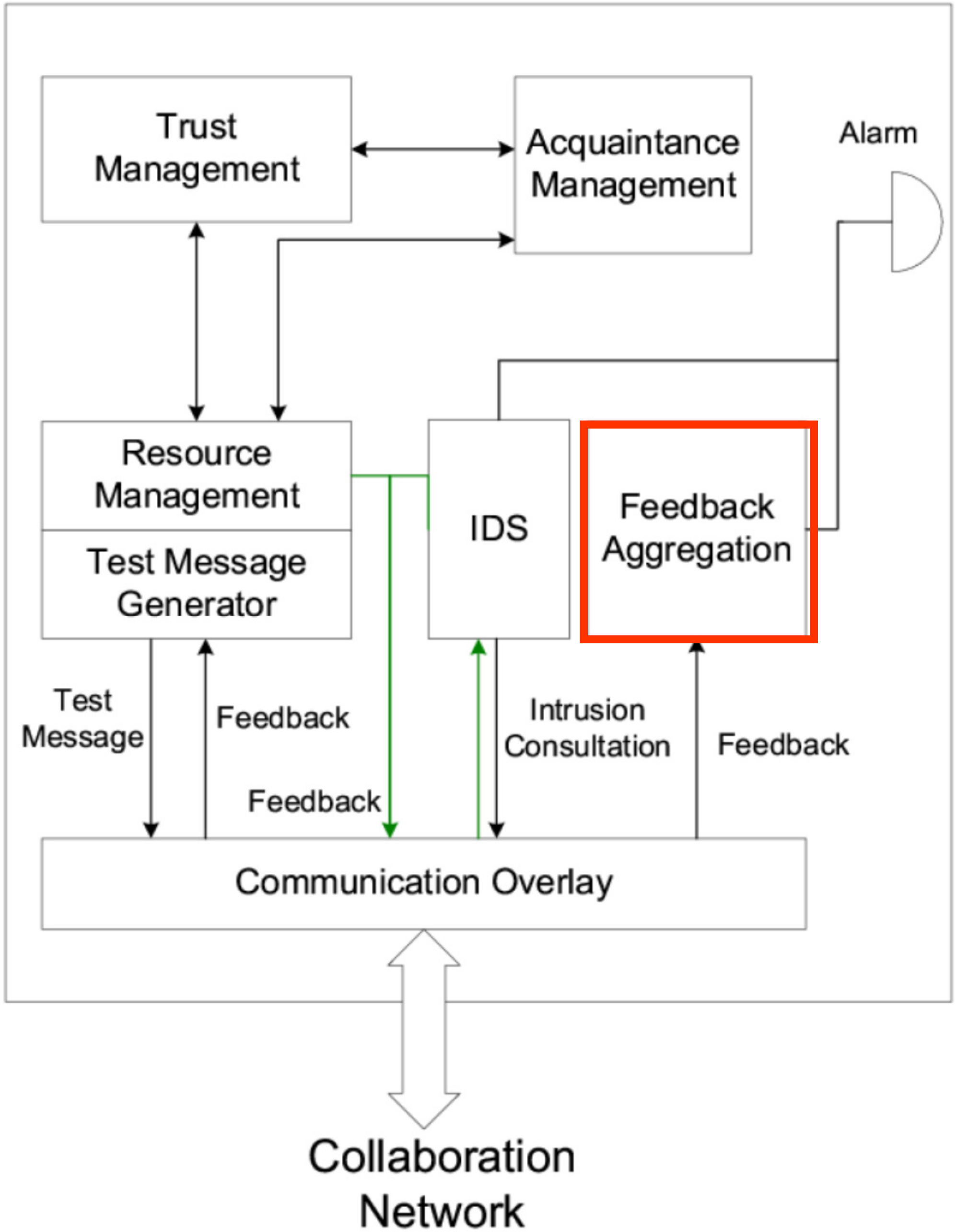
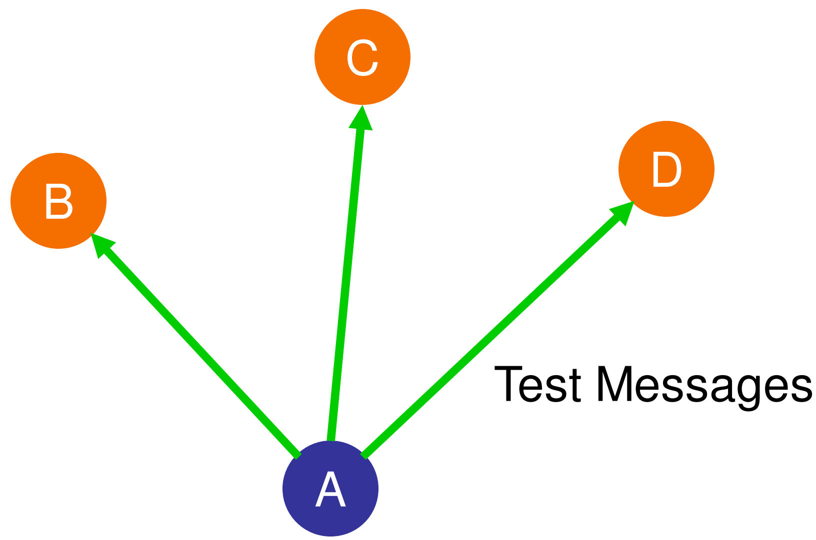
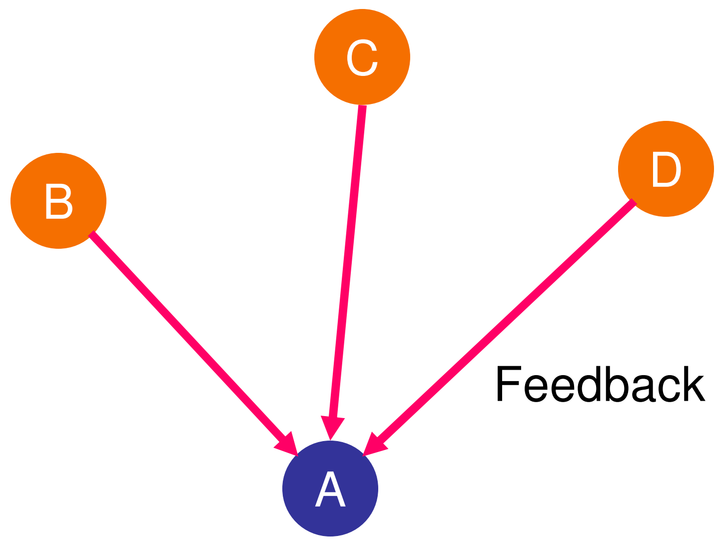


Figure 2. CIDN Architecture

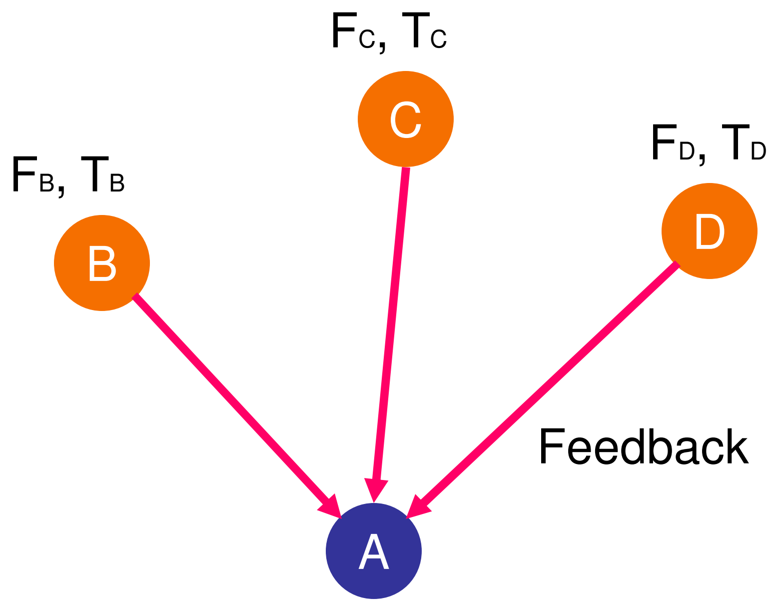
A Study Case



A Study Case

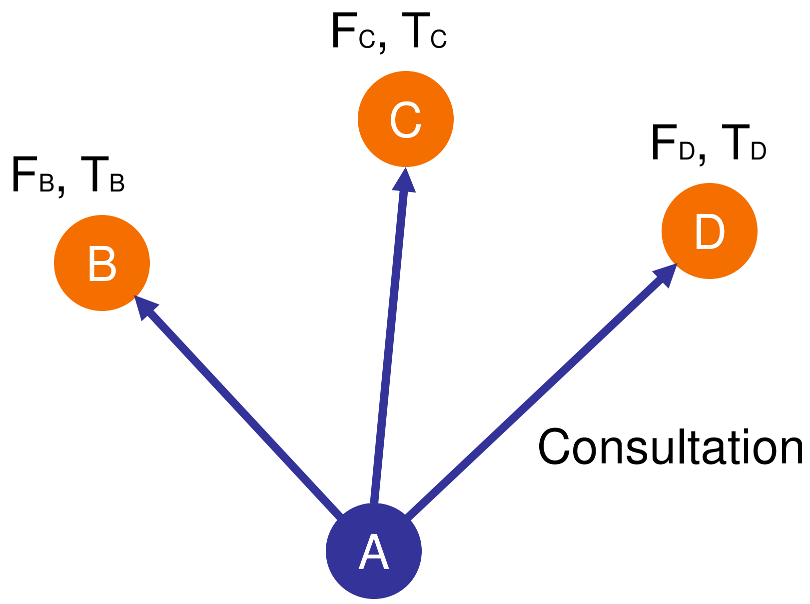


A Study Case

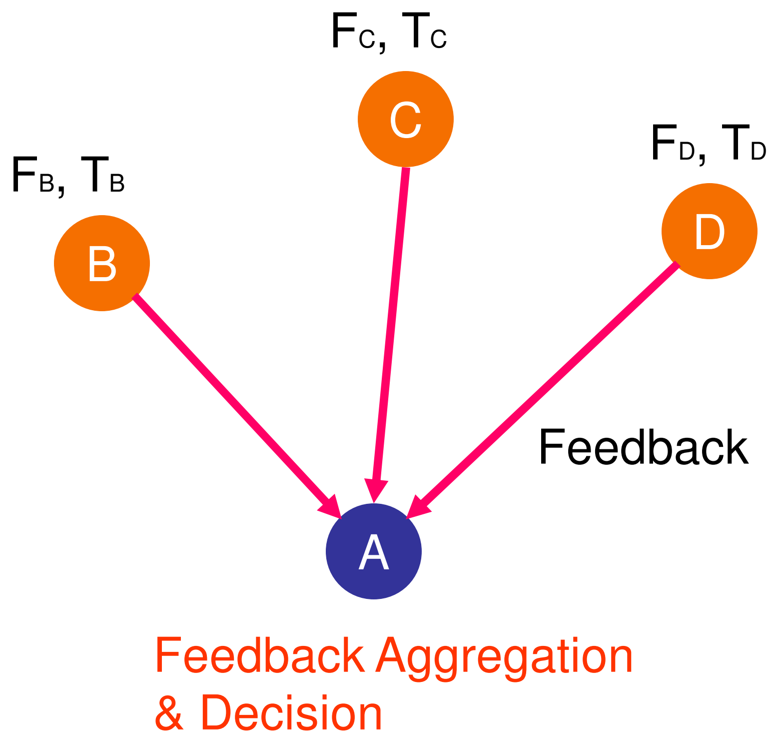


$F = P[Y=1|X=0]$
$T = P[Y=1|X=1]$

A Study Case



A Study Case



F_C, T_C
C
F_B, T_B
B
F_D, T_D
D
Feedback
A
Feedback Aggregation
& Decision

Problem Statement

- **Input:**
  - A number of n collaborators
  - The detection history of each collaborator
  - Prior probability of intrusions
  - Current feedback from each collaborator
  - The cost of false positive, false negative
- **Output:**
  - Final decision (yes/no)
- **Goal:**
  - Minimize expected cost of false decisions

Notations

n - Number of collaborators

π_1 - Prior probability of intrusion

$\{y_1,\ldots,y_n\}$ - Current feedback set

λ - Forgetting factor

$\{r^0_{k,1},\ldots,r^0_{k,m_k}\}$ - feedback history of node k for no intrusion test cases

$\{r^1_{k,1},\ldots,r^1_{k,n_k}\}$ - feedback history of node k for intrusion test cases

F_k - Probability that node k raises false alarm

T_k - Probability that node k raises true alarm

C_{fp} - Cost of a false positive decision

C_{fn} - Cost of a false negative decision

FP, TP Modeling

We use Beta distribution to model posterior probability of FP and TP

$$\mathcal{F}_k \sim \quad \mathrm{Beta}(x_k|\alpha_k^0, \beta_k^0) = \frac{\Gamma(\alpha_k^0+\beta_k^0)}{\Gamma(\alpha_k^0)\Gamma(\beta_i^0)} x_k^{\alpha_k^0 - 1}(1-x_k)^{\beta_k^0 - 1}$$

$$\mathcal{T}_k \sim \quad \mathrm{Beta}(y_k|\alpha_k^1, \beta_k^1) = \frac{\Gamma(\alpha_k^1+\beta_k^1)}{\Gamma(\alpha_k^1)\Gamma(\beta_i^1)} y_k^{\alpha_k^1 - 1}(1-y_k)^{\beta_k^1 - 1}$$

where,

$$\alpha_k^0 = \textstyle\sum_{j=1}^{u} \lambda^{t_{kj,}^0} r_{k,j}^0 \quad \beta_k^0 = \displaystyle\sum_{j=1}^{u} \lambda^{t_{k,j}^0}(1 - r_{k,j}^0)$$

$$\alpha_k^1 = \textstyle\sum_{j=1}^{v} \lambda^{t_{k,j}^1} r_{k,j}^1 \quad \beta_k^1 = \displaystyle\sum_{j=1}^{v} \lambda^{t_{k,j}^1}(1 - r_{k,j}^1)$$

Recursive Expression

$$
\begin{aligned}
\alpha_k^l(t_j) &= \lambda^{(t_{k,j}^l - t_{k,j-1}^l)} \alpha_k^l(t_{k,j-1}^l) + r_{k,j}^l \\
\beta_k^l(t_j) &= \lambda^{(t_{k,j}^l - t_{k,j-1}^l)} \beta_k^l(t_{k,j-1}^l) + r_{k,j}^l
\end{aligned}
$$

No need to keep all the history of all collaborators

Aggregation

$\mathbb{P}[X = 1 | \mathbf{Y} = \mathbf{y}]$

Aggregation

$$\mathbb{P}[X = 1|\mathbf{Y} = \mathbf{y}]$$

$$= \frac{\mathbb{P}[\mathbf{Y} = \mathbf{y}|X = 1]\mathbb{P}[X = 1]}{\mathbb{P}[\mathbf{Y} = \mathbf{y}|X = 1]\mathbb{P}[X = 1] + \mathbb{P}[\mathbf{Y} = \mathbf{y}|X = 0]\mathbb{P}[X = 0]}$$

Aggregation

$$\mathbb{P}[X = 1|\mathbf{Y} = \mathbf{y}]$$

$$= \frac{\mathbb{P}[\mathbf{Y} = \mathbf{y}|X = 1]\mathbb{P}[X = 1]}{\mathbb{P}[\mathbf{Y} = \mathbf{y}|X = 1]\mathbb{P}[X = 1] + \mathbb{P}[\mathbf{Y} = \mathbf{y}|X = 0]\mathbb{P}[X = 0]}$$

$$= \frac{\pi_1 \prod_{k=1}^{|\mathcal{A}|} T_k^{\mathbf{y}_k}(1 - T_k)^{1-\mathbf{y}_k}}{\pi_1 \prod_{k=1}^{|\mathcal{A}|} T_k^{\mathbf{y}_k}(1 - T_k)^{1-\mathbf{y}_k} + \pi_0 \prod_{k=1}^{|\mathcal{A}|} F_k^{\mathbf{y}_k}(1 - F_k)^{1-\mathbf{y}_i}}$$

Aggregation

$$\mathbb{P}[X = 1|\mathbf{Y} = \mathbf{y}]$$

$$= \frac{\mathbb{P}[\mathbf{Y} = \mathbf{y}|X = 1]\mathbb{P}[X = 1]}{\mathbb{P}[\mathbf{Y} = \mathbf{y}|X = 1]\mathbb{P}[X = 1] + \mathbb{P}[\mathbf{Y} = \mathbf{y}|X = 0]\mathbb{P}[X = 0]}$$

$$= \frac{\pi_1 \prod_{k=1}^{|\mathcal{A}|} T_k^{\mathbf{y}_k}(1 - T_k)^{1-\mathbf{y}_k}}{\pi_1 \prod_{k=1}^{|\mathcal{A}|} T_k^{\mathbf{y}_k}(1 - T_k)^{1-\mathbf{y}_k} + \pi_0 \prod_{k=1}^{|\mathcal{A}|} F_k^{\mathbf{y}_k}(1 - F_k)^{1-\mathbf{y}_i}}$$

Let P = $\mathbb{P}[X = 1|\mathbf{Y} = \mathbf{y}]$

The density function of P is denoted by $f_P(p)$

Decision

We model the cost of false decisions

$$
\begin{aligned}
& R(\delta) \\
= & \int_0^1 (C_{fp}(1-x)\delta + C_{fn}x(1-\delta)) f_P(x)dx \\
= & \; C_{fn}\mathbb{E}[P] + \delta(C_{fp} - (C_{fp}+C_{fn})\mathbb{E}[P])
\end{aligned}
$$

where

$\delta = 1$ Raise an intrusion alarm

$\delta = 0$ No alarm

Decision

$$\delta = \begin{cases} 1 \ \text{(Alarm)} & \text{if } \mathbb{E}[P] \geq \tau, \\ 0 \ \text{(No alarm)} & \text{otherwise.} \end{cases}$$

where $\tau = \frac{C_{fp}}{C_{fp} + C_{fn}}$

Gaussian Approximation

We need to calculate E[P] to make a decision

$$P = \frac{\pi_1 \prod_{k=1}^{|\mathcal{A}|} T_k^{\mathbf{y}_k} (1 - T_k)^{1-\mathbf{y}_k}}{\pi_1 \prod_{k=1}^{|\mathcal{A}|} T_k^{\mathbf{y}_k} (1 - T_k)^{1-\mathbf{y}_k} + \pi_0 \prod_{k=1}^{|\mathcal{A}|} F_k^{\mathbf{y}_k} (1 - F_k)^{1-\mathbf{y}_i}}$$

Gaussian Approximation

We need to calculate E[P] to make a decision

$$P = \frac{\pi_1 \prod_{k=1}^{|\mathcal{A}|} T_k^{\mathbf{y}_k} (1 - T_k)^{1-\mathbf{y}_k}}{\pi_1 \prod_{k=1}^{|\mathcal{A}|} T_k^{\mathbf{y}_k} (1 - T_k)^{1-\mathbf{y}_k} + \pi_0 \prod_{k=1}^{|\mathcal{A}|} F_k^{\mathbf{y}_k} (1 - F_k)^{1-\mathbf{y}_i}}$$

When the number of samples is large enough, Beta distribution can be approximated by Gaussian distribution

$$\mathbb{E}[P] \approx \frac{1}{1 + \frac{\pi_0}{\pi_1} \prod_{k=1}^{|\mathcal{A}|} \frac{\alpha_k^1 + \beta_k^1}{\alpha_k^0 + \beta_k^0} (\frac{\alpha_k^0}{\alpha_k^1})^{\mathbf{y}_k} (\frac{\beta_k^0}{\beta_k^1})^{1-\mathbf{y}_k}}$$

Cost of Decision

$$R(\delta) = \begin{cases} C_{fp}(1 - \mathbb{E}[P]) & \text{if } \mathbb{E}[P] \geq \tau \\ C_{fn}\mathbb{E}[P] & \text{otherwise.} \end{cases}$$

Optimal Decision Algorithm

Algorithm 1 Optimal_Decision(U_g, $\mathcal{A}$)

Require: $U_g \geq 0 \vee \mathcal{A} \neq \emptyset$
Ensure: $\delta(U_g, \mathcal{A})$
  $U \Leftarrow \infty$ {U is the current cost.}
  $Q \Leftarrow \frac{\pi_0}{\pi_1}$ {Note that $\mathbb{E}[P] = \frac{1}{1+Q}$ from (11).}
  while $\mathcal{A} \neq \emptyset \wedge U > U_g$ **do**
    {More consultation if cost is higher than threshold U_g}
    $a \Leftarrow$ firstElementOf($\mathcal{A}$)
    $\mathcal{A} \Leftarrow \mathcal{A} \setminus a$
    $r \Leftarrow$ getFeedback(a){Receive feedback from acquaintance a}
    if $r = 0$ **then**
      $Q \Leftarrow Q \cdot \frac{1-F(a)}{1-T(a)}$
    else
      $Q \Leftarrow Q \cdot \frac{F(a)}{T(a)}$
    end if
    $U \Leftarrow \min\left(\frac{C_{fp}Q}{1+Q}, \frac{C_{fn}}{1+Q}\right)$ {Get the lower cost of the two possible decisions}
  end while
  if $\frac{1}{1+Q} > \frac{C_{fp}}{C_{fp}+C_{fn}}$ **then**
    Raise Alarm
  else
    No Alarm
  end if

Simulation Result

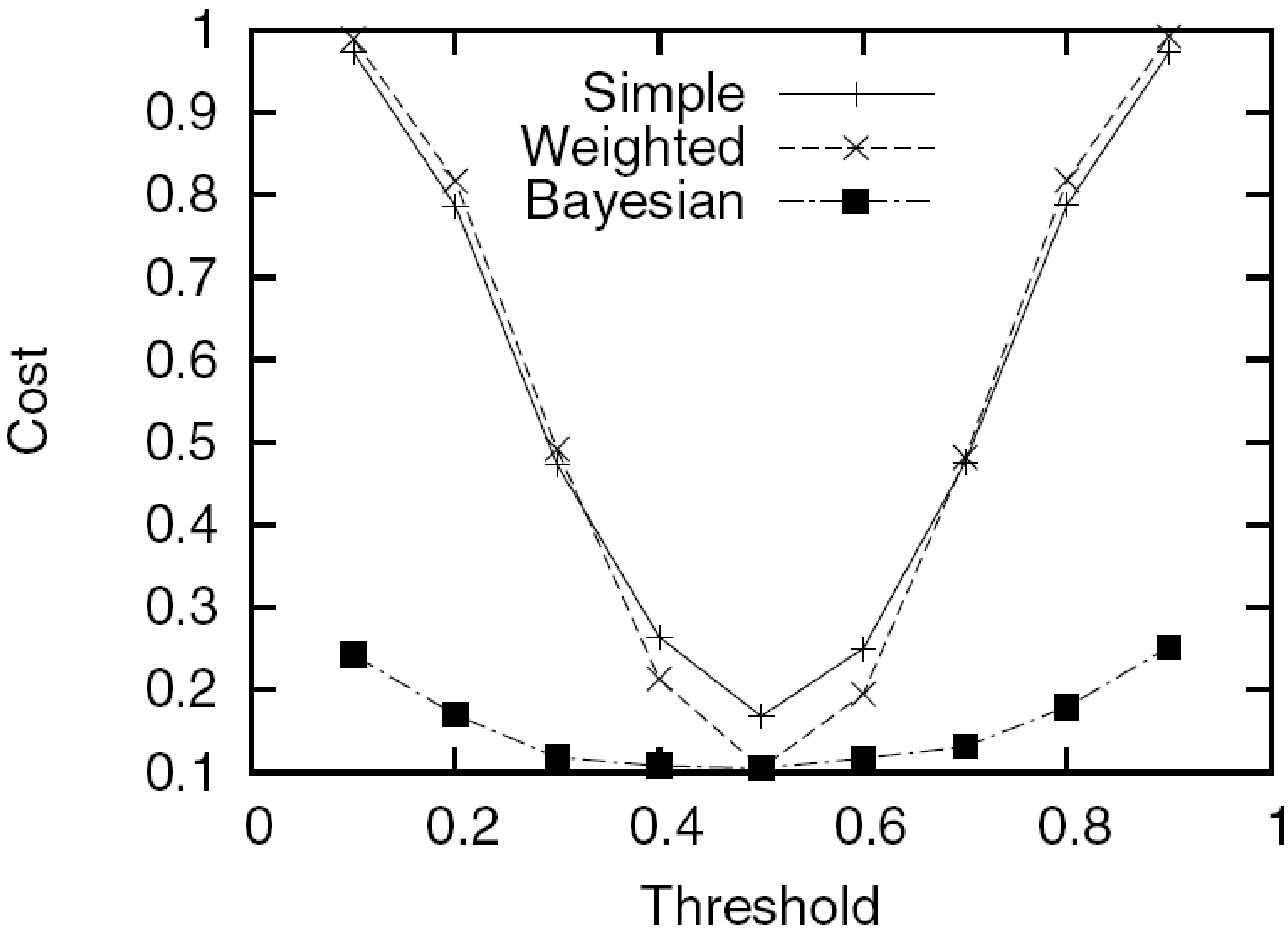


Figure 3. Comparison of cost using different aggregation techniques

Simulation Result

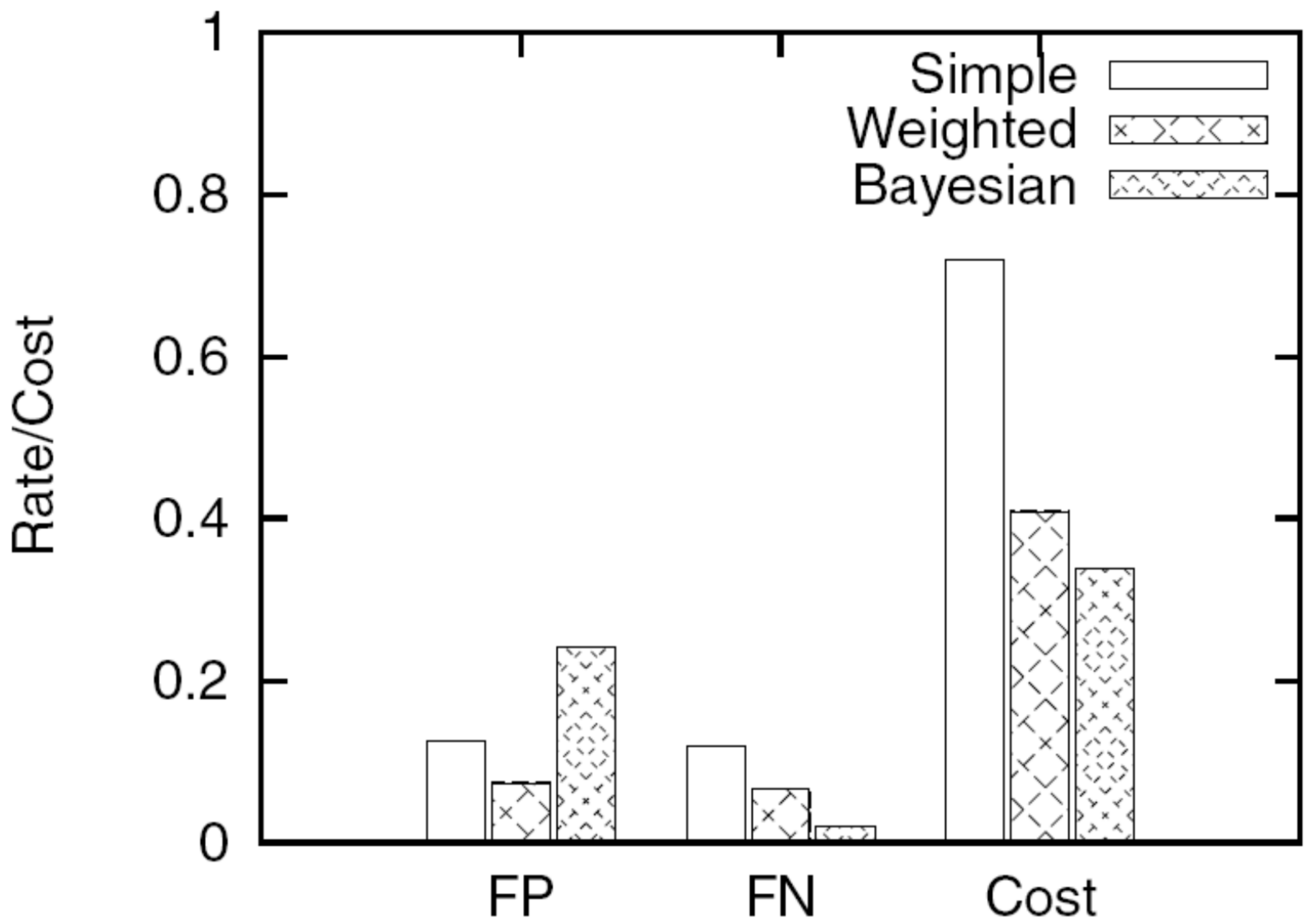


Figure 4. Comparison of FP, FN, and cost

Simulation Result

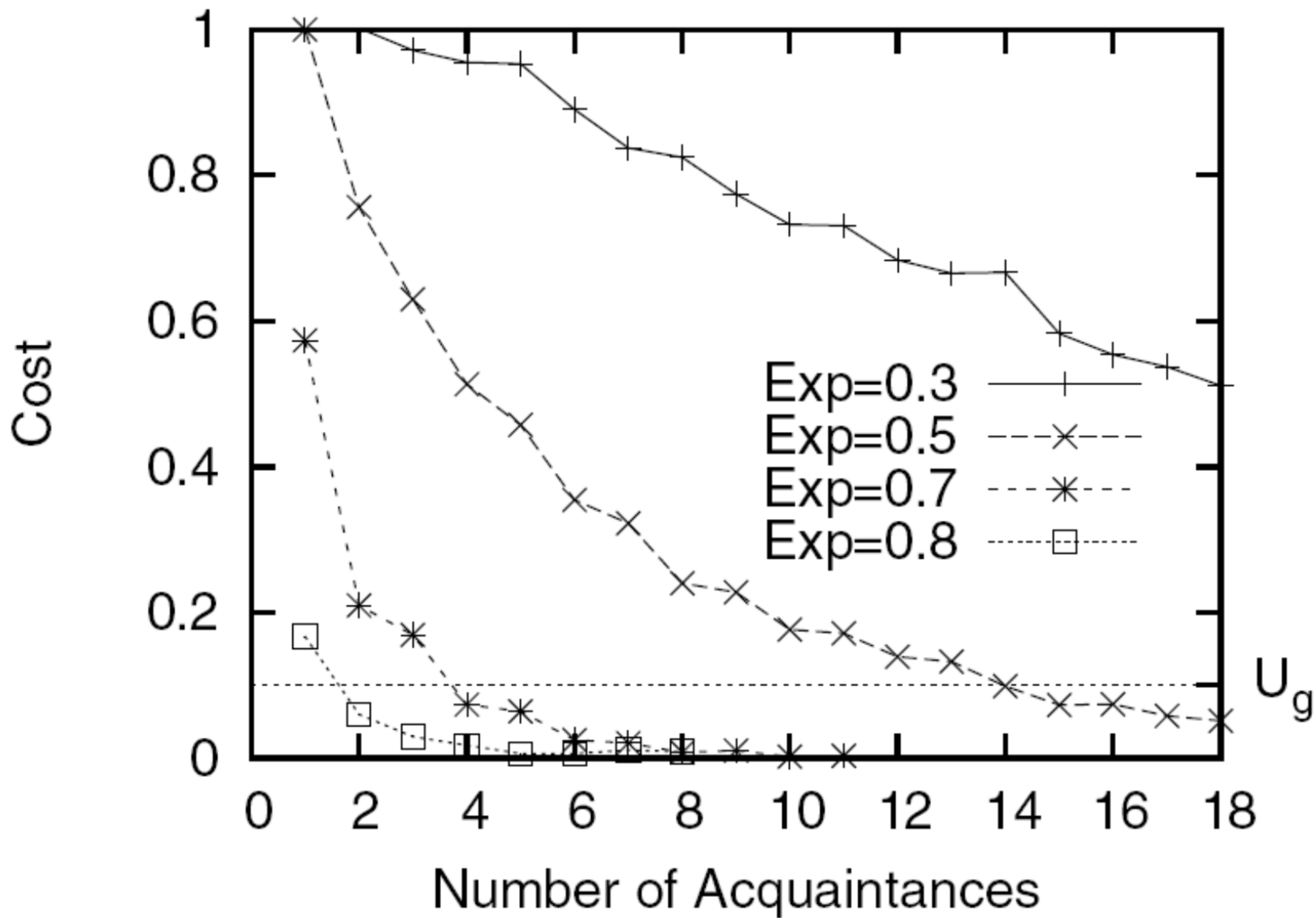


Figure 5. Average Cost vs. Number of Acquaintances Consulted

Conclusions and Future Work

- Framework of a distributed collaborative intrusion detection network
- A Bayesian aggregation and decision model to minimize expected cost
- Dynamic online aggregation and decision
- As our future work, we intent to implement and deploy our CIDN on real life open source IDSes

Questions