P3CA: Privacy Preserving Traffic Anomaly Detection for ISP Networks

Virajith Jalaparti, Shishir Nagaraja, Matthew Caesar, Nikita Borisov

University of Illinois at Urbana-Champaign

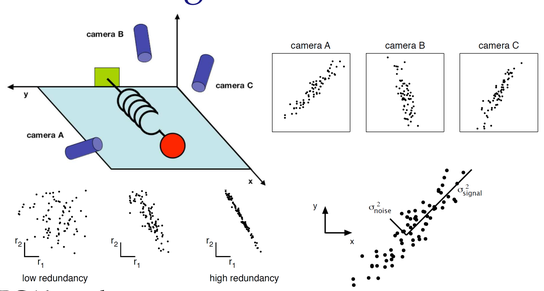
Introduction

- Anomalies in traffic patterns are a common occurrence in today's Internet.
  - Affects end-users: DoS, Botnets, etc.
  - Affects network performance: congestion on links
- Anomaly detection is important and challenging.
  - Used to prevent economic losses for both network operators and end-users.
  - Requires continuous in-network monitoring of networks in a scalable manner.
- Subspace analysis using Principal Components (PCA) has been proven to be quite effective.
- Cooperation among ISPs can lead to increased chance of detecting anomalies.
  - Larger number of vantage points and different "mixes" of traffic.
  - Requires parties involved to share details of their traffic: which leads to privacy concerns.

Contribution: Privacy Preserving PCA

- P3CA: A Secure Multiparty Cooperation-based semi-centralized variant of PCA.
- P3CA performs anomaly detection across ISPs in a privacy preserving manner while being:
  - Scalable: $O(\#observations)/party$.
  - Compatible with existing protocols.
  - Low in maintainance.
- Privacy Goals: Does not reveal details of:
  - Network Topology.
  - Workload/Traffic Information.
  - Monitoring Infrastructures.

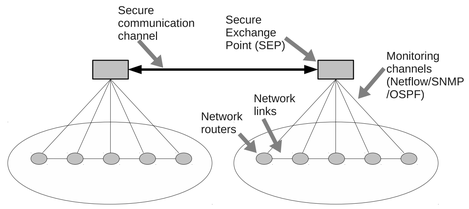
Background: PCA



PCA's goals:

- *Minimize the redundancy in measurements.*
- *Maximize the signal along the basis.*

System Architecture



Protocol Overview

1. All ISPs select two special representatives:
   - Alice, the coordinator who collects data encrypted by Bob's public key.
   - Bob, the keyholder who performs operations on blinded data.
2. All ISPs execute P3CA,to generate a set of principal components, $P = [p_1|p_2|\dots|p_n]$. P3CA uses:
   - A variant of the Power Method made privacy preserving using:
     - *Private Vector Normalization.*
     - *Private Matrix-Vector Multiplication.*
     - *Private Number Comparison.*
   - Paillier Encryption Scheme for homomorphic encryption, which provides:
     - $Enc(v_1 + v_2) = Enc(v_1) \oplus Enc(v_2) = Enc(v_1) * Enc(v_2)$.
     - $Enc(p \times v) = p \otimes Enc(v) = Enc(v)^p$.
3. Each ISP k uses P to find the the residual traffic matrix, $R_k = (I - PP^T)T_k$ (T_k: traffic matrix of ISP k). and uses the Q-statistic to detect the anomalies in its network.

Power Method

- Iterative procedure used to calculate principal components of a matrix.

Input: $t \times m$ matrix T
Input: τ is a convergence parameter
Output: Top k-eigenpairs of TT^T namely, $(\lambda_1, x_1), \dots, (\lambda_n, x_n)$
foreach *Eigenpair* $(\lambda_{q\leq k}, x_{q\leq k})$ *to be calculated* **do**
  $\delta \leftarrow 1$; $v \leftarrow random\_vector()$; $S \leftarrow t \times t$ zeros;
  while $\delta \geq \tau|\lambda_q|$ **do**
    $n = \frac{v}{||v||}$;
    $u = T^T n$;
    $w = Tu$;
    $v = w - Sn$;
    $\lambda_q = n^T v$;
    $\delta = v - n\lambda_q$;
  end
  $x_q = v$;
  $S = S + \lambda_q(n^T * n)$;
end

Private Vector Normalization

- Goal: Find $Enc(\hat{v})$ given $Enc(\vec{v})$, $\hat{v} = \frac{\vec{v}}{||\vec{v}||}$.
- Preserve privacy of:
  - Magnitude of $\vec{v}$:
    Blinding: Multiply with a random number

$$Enc(\vec{v'}) = r \otimes Enc(\vec{v})$$

  - Direction of $\vec{v}$
    Blinding: Rotate $\vec{v}$ by a random angle

$$R_i = \begin{bmatrix} & & \text{col. } \mathbf{p} & & \text{col. } \mathbf{q} & \\ & & \downarrow & & \downarrow & \\ & 1 \dots & 0 & \dots & 0 & \dots\ 0 \\ & 0 \dots & 0 & \dots & 0 & \dots\ 0 \\ \mathbf{p}\rightarrow & 0 \dots & cos(\theta_i) & \dots & -sin(\theta_i) & \dots\ 0 \\ & \vdots & & \ddots & & \vdots \\ \mathbf{q}\rightarrow & \dots & sin(\theta_i) & \dots & cos(\theta_i) & \dots\ 0 \\ & 0 \dots & 0 & \dots & 0 & \dots\ 0 \\ & 0 \dots & 0 & \dots & 0 & \dots\ 1 \end{bmatrix}$$

$$Enc(\vec{v_r}) = R_c \otimes R_{c-1} \otimes \dots \otimes R_1 \otimes Enc(\vec{v'})$$

- Alice sends $Enc(\vec{v_r})$ to Bob, who returns $Enc(\hat{v_r})$.
- Alice recovers $Enc(\hat{v})$ by $(R_c \otimes R_{c-1} \otimes \dots \otimes R_1)^T \otimes Enc(\hat{v_r})$

Private Matrix-Vector Multiplications

- Goal: Find $Enc(R)$ where $R = T \cdot (T^T v)$, $T = [T_1|T_2|\dots|T_p]$ and ISP i owns T_i.
- $[v'_1, v'_2, \dots v'_k] = (T^T v)$ computed in a a distributed manner.
- ISP i computes $(v_a)_i = [(T_i)_1 \cdot v'_{i_1}, (T_i)_2 \cdot v'_{i_2}, \dots (T_i)_m \cdot v'_{i_m}]$ and sends $Enc((v_a)_i)$ to Alice.
- *Blinding*: Alice adds random vectors $r_i \in \mathbb{R}^m$ to $Enc((v_a)_i)$ and uses Bob's help to calculate $Enc(\sum_{i=1}^{n}(v_a)_i)$.

Private Number Comparison

- Goal: Find if $a \geq b$ given $Enc(a)$ and $Enc(b)$
- *Blinding*: Alice sends $Enc(a + r - b)$ to Bob where $r \in \mathbb{R}$ is a random number.
- Bob decrypts the result and sends the plaintext c to Alice.
- Alice can now compare c with r to evaluate whether $a > b$ is true.

Experimental Results

Performance and Scalability

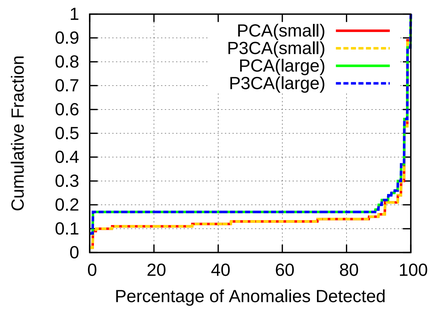
$\#links$	$\#bins$	Party i	Alice
20	2	0.525	2.1826
40	4	1.376	6.4720
80	8	3.529	14.1341
160	16	52.999	194.175
320	32	194.126 (.05h)	637.649 (.17h)

P3CA: Computation time (in sec) at each party and at Alice (the coordinator)

Percent of time	Operation
39.6	multiplying cipher and plain texts
36.6	adding cipher texts
18.2	decryption
16.5	private vector normalization
82.52	private matrix-vector multiplication and subtractions

P3CA: Microbenchmarks

Comparison with PCA



Future Directions

- Address the limitations of P3CA:
  - Cannot be used online: needed in reality.
  - Assumes honest-but-curious: in practice malicious ISPs can exist.
- Using an incremental PCA mechanism to deal with small changes in data.
- Framework to actively detect, localize and deal with anomalies.
- Integration with techniques that deal with corruption of data used for PCA.