Technical Report TSC/FPG/20032014: Performance Analysis of the Least Squares Disclosure Attack with Time-Varying User Behavior

Fernando Pérez-González Carmela Troncoso Simon Oya

December 3, 2015

In this report, we provide a performance analysis of the Least Squares Disclosure Attack (LSDA) described in [1] when the user behavior is not static. In order to account for dynamic behavior, we assume that the sending frequencies and user profiles in each round, $\{f_i^r\}$ and $\{p_{j,i}^r\}$ respectively, correspond to realizations of the random processes $\{F_i^r\}$ and $\{P_{j,i}^r\}$. In order to keep the analysis tractable, we will limit our derivations to the case where $\{F_i^r\}$ and $\{P_{j,i}^r\}$ are wide-sense stationary, and the input process $\{X_i^r\}$ is ergodic. The basic notation used in this document is described in Sect. III of [1].

1 Unbiased estimator of the average profile

We now prove that LSDA is an *unbiased estimator* of the *average sending profile* of the users in the system when the process modeling the variations in the profiles, $\{P_{j,i}^r\}$, is wide-sense stationary.

We start by noting that, in this dynamic scenario, the output process can be modeled as the sum of N multinomials:

$$\{Y_1^r, Y_2^r, \cdots, Y_N^r|\mathbf{U}, \mathbf{P}\} \sim \sum_{i=1}^{N} \text{Multi}\left(X_i^r, \{P_{1,i}^r, P_{2,i}^r, \cdots, P_{N,i}^r\}\right) \tag{1}$$

Let $\mathbf{U}^r$ be the $\rho \times N$ matrix which contains in its r-th row $[X_1^r, \cdots, X_N^r]$ and zeros in all other positions, and note that $\sum_{r=1}^{\rho} \mathbf{U}^r = \mathbf{U}$. We can then write

$$\mathrm{E}\left\{\mathbf{Y}_j|\mathbf{U}\right\} = \mathrm{E}\left\{\sum_{r=1}^{\rho} \mathbf{U}^r \mathbf{P}_j^r|\mathbf{U}\right\} = \sum_{r=1}^{\rho} \mathbf{U}^r \mathrm{E}\left\{\mathbf{P}_j\right\} = \mathbf{U} \cdot \mathrm{E}\left\{\mathbf{P}_j\right\}$$

. This result allows us to show that

$$\mathrm{E}\left\{\hat{\mathbf{p}}_j\right\} = \mathrm{E}\left\{\mathrm{E}\left\{\hat{\mathbf{p}}_j|\mathbf{U}\right\}\right\} = \mathrm{E}\left\{(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathrm{E}\left\{\mathbf{Y}_j|\mathbf{U}\right\}\right\} = \mathrm{E}\left\{(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{U} \cdot \mathrm{E}\left\{\mathbf{P}_j\right\}\right\} = \mathrm{E}\left\{\mathbf{P}_j\right\} \tag{2}$$

which concludes the proof.

2 Performance analysis with dynamic user behavior

We aim at finding an expression for the Mean Squared Error per user, $\mathrm{MSE}_i \doteq \sum_{j=1}^{N} \mathrm{Var}\left\{\hat{p}_{j,i}\right\}$, when using the LSDA estimator in a threshold mix scenario with dynamic sending frequencies and user profiles. We limit this performance analysis to the scenario where the processes modeling this behavior are wide-sense stationary. Following the derivations in [1], we start by using (2) to get $\mathrm{Var}\left\{\mathrm{E}\left\{\hat{p}_{j,i}|\mathbf{U}\right\}\right\} = 0$ and $\mathrm{Cov}\left\{\mathrm{E}\left\{\hat{p}_{j,i}|\mathbf{U}\right\}, \mathrm{E}\left\{\hat{p}_{j,k}|\mathbf{U}\right\}\right\} = 0$, so that we can write the covariance matrix of $\hat{\mathbf{p}}_j$ as

$$\mathbf{\Sigma}_{\hat{\mathbf{p}}_\mathbf{j}} = \mathrm{E}\left\{\mathbf{\Sigma}_{\hat{\mathbf{p}}_\mathbf{j}|\mathbf{U}}\right\} = \mathrm{E}\left\{(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\right\} \tag{3}$$

For an ergodic input process, we can write $\lim_{\rho\to\infty} \mathbf{U}^T\mathbf{U}/\rho \to \mathbf{R}_x$ where the m,n-th element of $\mathbf{R}_x$ is $\mathrm{E}\{X_m X_n\}$. Therefore, if the number of observed rounds ρ is large, we can approximate (3) as

$$\mathbf{\Sigma}_{\hat{\mathbf{p}}_\mathbf{j}} \approx \frac{1}{\rho^2}\mathbf{R}_x^{-1}\mathrm{E}\left\{\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right\}\mathbf{R}_x^{-1} \tag{4}$$

The next step would be computing the terms $\mathbf{R}_x^{-1}$ and $\mathrm{E}\left\{\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right\}$. For tractability issues, we carry out these derivations by considering the cases of dynamic sending frequencies and user profiles separately, and then we conclude by giving the intuition of what happens when both vary simultaneously.

2.1 Derivation of MSE_i for dynamic profiles

We first assess the performance of the estimator in the scenario where the user profiles $p_{j,i}^r$ change between rounds and the sending frequencies are static, i.e., $f_i^r = f_i^s$ for all r, s. Here, the autocorrelation matrix of the input process and its inverse can be found on [1], where it is shown that

$$\mathbf{R}_x^{-1} = \frac{1}{t}\left[{\mathbf{\Lambda}_\mathbf{F}}^{-1} - \left(1-\frac{1}{t}\right)\mathbf{1}_{N\times N}\right]. \tag{5}$$

where $\mathbf{\Lambda}_\mathbf{F} \doteq \mathrm{diag}\{\mathbf{f}\}$, $\mathbf{f} \doteq [f_1, \cdots, f_N]^T$.

In order to compute the remaining term in (4), $\mathrm{E}\left\{\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right\}$, we use the law of total variance together with (1), expanding the terms in $\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}$ as

$$\begin{aligned}
\mathrm{Var}\left\{Y_j^r|\mathbf{U}\right\} &= \mathrm{Var}\left\{\mathrm{E}\left\{Y_j^r|\mathbf{U},\mathbf{P}_j\right\}|\mathbf{U}\right\} + \mathrm{E}\left\{\mathrm{Var}\left\{Y_j^r|\mathbf{U},\mathbf{P}_j\right\}|\mathbf{U}\right\} \\
&= \mathrm{Var}\left\{\textstyle\sum_{i=1}^N X_i^r P_{j,i}^r|\mathbf{U}\right\} + \mathrm{E}\left\{\textstyle\sum_{i=1}^N X_i^r P_{j,i}^r(1-P_{j,i}^r)|\mathbf{U}\right\} \\
&= \textstyle\sum_{i=1}^N (X_i^r)^2\mathrm{Var}\{P_{j,i}\} + \sum_{i=1}^N X_i^r\mathrm{E}\{P_{j,i}(1-P_{j,i})\}
\end{aligned} \tag{6}$$

$$\begin{aligned}
\mathrm{Cov}\left\{Y_j^r, Y_j^s|\mathbf{U}\right\} &= \mathrm{Cov}\left\{\mathrm{E}\left\{Y_j^r|\mathbf{U},\mathbf{P}_j\right\}, \mathrm{E}\left\{Y_j^s|\mathbf{U},\mathbf{P}_j\right\}|\mathbf{U}\right\} + \mathrm{E}\left\{\mathrm{Cov}\left\{Y_j^r, Y_j^s|\mathbf{U},\mathbf{P}_j\right\}|\mathbf{U}\right\} \\
&= \mathrm{Cov}\left\{\textstyle\sum_{i=1}^N X_i^r P_{j,i}^r, \sum_{k=1}^N X_k^s P_{j,k}^s|\mathbf{U}\right\} \\
&= \textstyle\sum_{i=1}^N\sum_{k=1}^N X_i^r X_k^s \mathrm{Cov}\left\{P_{j,i}^r, P_{j,k}^s\right\} \qquad r \neq s
\end{aligned} \tag{7}$$

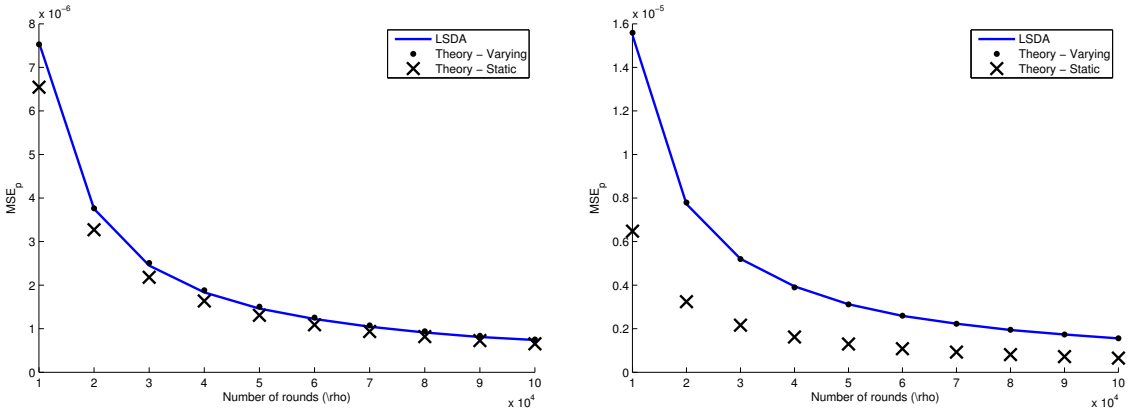
We disregard the dependence between the transition probabilities of different rounds $\mathrm{Cov}\left\{P_{j,i}^r, P_{j,k}^s\right\} \approx 0$ for $r \neq s$ since this term is small compared to $\mathrm{Var}\left\{P_{j,i}^r\right\}$. Then, the m,n-th element of $\mathrm{E}\left\{\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right\}$ is

$$\left(\mathrm{E}\left\{\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right\}\right)_{m,n} = \rho\sum_{i=1}^N \mathrm{E}\{X_m X_n X_i\}\,\mathrm{E}\{P_{j,i}(1-P_{j,i})\} + \rho\sum_{i=1}^N \mathrm{E}\left\{X_m X_n X_i^2\right\}\mathrm{Var}\{P_{j,i}\} \tag{8}$$

This term can be written in matricial way as

$$\begin{aligned}
\tfrac{1}{\rho}\mathrm{E}\left\{\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right\} &= \mathbf{\Lambda}_\mathbf{F}\left(\eta_j t^{(3)}\mathbf{1}_{N\times N} + \mathbf{S}_j\mathbf{1}_{N\times N}t^{(2)} + \mathbf{1}_{N\times N}\mathbf{S}_j t^{(2)}\right)\mathbf{\Lambda}_\mathbf{F} \\
&+ \left(\eta_j t^{(2)}\mathbf{I}_{N\times N} + t\mathbf{S}_j\right)\mathbf{\Lambda}_\mathbf{F} \\
&+ \mathbf{\Lambda}_\mathbf{F}\left(\tilde{\eta}_j t^{(4)}\mathbf{1}_{N\times N} + 2t^{(3)}\left(\mathbf{\Lambda}_\mathbf{F}\tilde{\mathbf{S}}_j\mathbf{1}_{N\times N} + \mathbf{1}_{N\times N}\tilde{\mathbf{S}}_j\mathbf{\Lambda}_\mathbf{F}\right)\right)\mathbf{\Lambda}_\mathbf{F} \\
&+ \left(\tilde{\eta}_j t^{(3)}\mathbf{I}_{N\times N} + 4t^{(2)}\tilde{\mathbf{S}}_j\mathbf{\Lambda}_\mathbf{F}\right)\mathbf{\Lambda}_\mathbf{F}
\end{aligned} \tag{9}$$

where $t^{(k)} \doteq t(t-1)\cdots(t-k+1)$, $\eta_j \doteq \sum_{i=1}^N f_i s_{j,i}$, $\mathbf{S}_j \doteq \mathrm{diag}\{s_{j,1}, \cdots, s_{j,N}\}$, $\tilde{\eta}_j \doteq \sum_{i=1}^N f_i^2 \tilde{s}_{j,i}$ and $\tilde{\mathbf{S}}_j \doteq \mathrm{diag}\{\tilde{s}_{j,1}, \cdots, \tilde{s}_{j,N}\}$; with $s_{j,i} \doteq \mathrm{E}\{P_{j,i}(1-P_{j,i})\} + \mathrm{Var}\{P_{j,i}\} = \mathrm{E}\{P_{j,i}\}(1-\mathrm{E}\{P_{j,i}\})$ and $\tilde{s}_{j,i} \doteq \mathrm{Var}\{P_{j,i}\}$.



(a) $N = 10$, $t = 10$ (b) $N = 10$, $t = 100$

(c) $N = 100$, $t = 10$ (d) $N = 100$, $t = 100$

Figure 1: MSE_p evolution with the number of rounds ρ, in an scenario where the sending profiles vary in each round ($N = 10$, 100, $n_f = 25$, $f_i = 1/N$, $t = 10$, 100).

Plugging this term into (4) together with (5), we obtain an expression for the variance in the estimation of a single transition probability, $\mathrm{Var}\{\hat{p}_{j,i}\}$. Adding those variances along j, we finally get

$$\begin{aligned} \mathtt{MSE}_i &\approx \frac{1}{\rho}\left\{(f_i^{-1}-1)\left(1-\frac{1}{t}\right)\bar{\mu}+\frac{f_i^{-1}}{t}\mu_i\right\} \\ &+ \frac{1}{\rho}\left(1-\frac{1}{t}\right)\left\{\left[(t-2)f_i^{-1}-(t-3)\right]\bar{\sigma}+4\sigma_i f_i(f_i^{-1}-1)\right\} \end{aligned} \tag{10}$$

where $\mu_i \doteq 1 - \sum_j \mathrm{E}\{P_{j,i}\}^2$, $\bar{\mu} \doteq \sum_i f_i\mu_i$, $\sigma_i \doteq \sum_j \mathrm{Var}\{P_{j,i}\}$ and $\bar{\sigma} \doteq \sum_i f_i^2\sigma_i$.

This formula shows that time-varying sending profiles slow the attacker's attempt at estimating the average sending profiles, compared to the case where the sending profiles remain static between rounds. Also, the MSE increase introduced by dynamic sending profiles, represented by the second summand in (10), grows with t but decreases fast as the sending frequencies decrease (i.e., as N grows), since the parameter $\bar{\sigma}$ depends on f_i^2. Therefore, the increase in MSE introduced by time-varying sending profiles is normally small (e.g., if $N > 100$).

Figure 1 shows the average MSE obtained by repeating an experiment where sending profiles are chosen in each round as a realization of a random process, for two different values of the number of users in the system N and batch size t. We see that our theoretical approximation (10) closely models the average MSE of the attack in this scenario. The figure also shows how large values of t increase the MSE (Figs. 1b and 1d), but increasing the number of users quickly diminishes the effect of time-varying profiles (Figs. 1c and 1d).

2.2 Derivation of MSE_i for dynamic sending frequencies

We now compute the MSE_i assuming that user profiles are static, i.e., $p^r_{j,i} = p^s_{j,i}$ for all r, s, but sending frequencies vary between rounds. To make the derivations easier, we will assume that $\text{E}\{F_i\}^{-1} \gg 1$. In this scenario, the autocorrelation matrix of the input process can be written as

$$\mathbf{R}_x = t\left[\text{E}\{\mathbf{\Lambda_F}\} + (t-1)\left(\text{E}\{\mathbf{f}\}\,\text{E}\{\mathbf{f}\}^T + \mathbf{\Sigma_f}\right)\right] \tag{11}$$

where $(\mathbf{\Sigma_f})_{m,n} = \text{Cov}\{F_m, F_n\}$. We assume that we can disregard the terms $\text{Cov}\{F_m, F_n\}$ when $m \neq n$ because they are small compared to the diagonal terms $\text{Var}\{F_m\}$, and therefore consider $\mathbf{\Sigma_f} \approx \tilde{\mathbf{\Sigma}}_\mathbf{f} \doteq \text{diag}\{\text{Var}\{F_1\}, \cdots, \text{Var}\{F_N\}\}$. With this approximation, using Sherman-Morrison formula [2] yields

$$\mathbf{R}_x^{-1} = \text{E}\{\mathbf{\Lambda_F}\}^{-1}\mathbf{A}^{-1}\left(\mathbf{A} - \frac{t^{(2)}}{1 + t^{(2)}\text{tr}\{\mathbf{A}^{-1}\}}\right)\mathbf{A}^{-1}\text{E}\{\mathbf{\Lambda_F}\}^{-1} \tag{12}$$

where $\mathbf{A} \doteq t\left[\text{E}\{\mathbf{\Lambda_F}\}^{-1} + (t-1)\text{E}\{\mathbf{\Lambda_F}\}^{-1}\tilde{\mathbf{\Sigma}}_\mathbf{f}\text{E}\{\mathbf{\Lambda_F}\}^{-1}\right]$ and $\text{tr}\{\cdot\}$ denotes the trace operation.

On the other hand, $\text{E}\left\{\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right\}$ can be written as

$$\begin{aligned}\frac{1}{\rho}\text{E}\left\{\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right\} &= t^{(3)}\sum_{k=1}^{N} s_{j,k}\text{E}\left\{F_k\mathbf{f}\cdot\mathbf{f}^T\right\} + t^{(2)}\left(\mathbf{S}_j\text{E}\left\{\mathbf{f}\cdot\mathbf{f}^T\right\} + \mathbf{S}_j\text{E}\left\{\mathbf{f}\cdot\mathbf{f}^T\right\}\right) \\ &+ t^{(2)}\sum_{k=1}^{N} s_{j,k}\text{E}\{F_k\mathbf{\Lambda_F}\} + t\mathbf{S}_j\text{E}\{\mathbf{\Lambda_F}\}\end{aligned} \tag{13}$$

where $s_{j,i} = p_{j,i}(1 - p_{j,i})$ and $\mathbf{S}_j \doteq \text{diag}\{s_{j,1}, \cdots, s_{j,N}\}$.

As before, we consider that $\text{Cov}\{F_m, F_n\} \approx 0$ when $m \neq n$ and, additionally, $\text{Cov}\{F_k, F_m F_n\} \approx 2\text{E}\{F_k\}\,\text{Var}\{F_k\}$ when $k = m = n$ and zero otherwise. With these approximations, (13) can be rewritten as

$$\begin{aligned}\frac{1}{\rho}\text{E}\left\{\mathbf{U}^T\mathbf{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right\} &\approx t^{(3)}\eta_j\left(\text{E}\{\mathbf{f}\}\,\text{E}\{\mathbf{f}\}^T + \tilde{\mathbf{\Sigma}}_\mathbf{f}\right) + 2t^{(3)}\text{E}\{\mathbf{\Lambda_F}\}\,\tilde{\mathbf{\Sigma}}_\mathbf{f} \\ &+ t^{(2)}\left[\mathbf{S}_j\left(\text{E}\{\mathbf{f}\}\,\text{E}\{\mathbf{f}\}^T + \tilde{\mathbf{\Sigma}}_\mathbf{f}\right) + \mathbf{S}_j\left(\text{E}\{\mathbf{f}\}\,\text{E}\{\mathbf{f}\}^T + \tilde{\mathbf{\Sigma}}_\mathbf{f}\right)\right] \\ &+ t^{(2)}\left(\eta_j\text{E}\{\mathbf{\Lambda_F}\} + \mathbf{S}_j\tilde{\mathbf{\Sigma}}_\mathbf{f}\right) + t\mathbf{S}_j\text{E}\{\mathbf{\Lambda_F}\}\end{aligned} \tag{14}$$

where $\eta_j \doteq \sum_{i=1}^{N}\text{E}\{F_i\}\,s_{j,i}$.

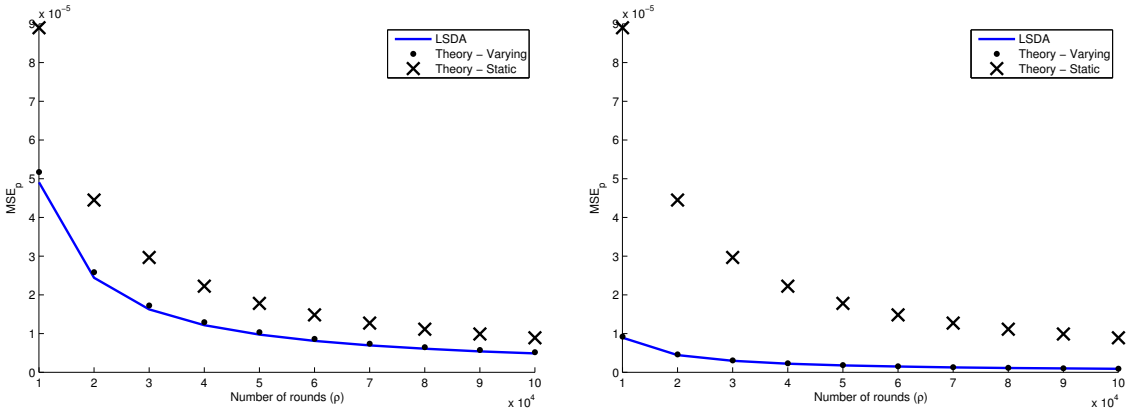
Plugging (14) and (12) into (4) we obtain an approximation for $\text{Var}\{\hat{p}_{j,i}\}$. Adding along j, we finally get

$$\text{MSE}_i \approx \frac{1}{\rho}\cdot\frac{1}{\text{E}\{F_i\}}\cdot\frac{2t^{(3)}\mu_i\text{Var}\{F_i\} + t^{(2)}\bar{\mu} + t\mu_i + \frac{\text{Var}\{F_i\}}{\text{E}\{F_i\}}\left(t^{(3)}\bar{\mu} + 3t^{(2)}\mu_i\right)}{\left(t + t^{(2)}\frac{\text{Var}\{F_i\}}{\text{E}\{F_i\}}\right)^2} \tag{15}$$

Using (15), it can be shown that increasing the variance of the sending frequencies in turn decreases the MSE. Intuitively, the larger the variance in the input process, the larger the probability that one user is going to dominate in a given round. Observations where a user dominates give the attacker very valuable information about users' behavior, which results in a better estimation of the profiles. This decrease in MSE its even more pronounced when the batch size t is large. This is shown in Fig. 2, where we plot the average MSE obtained through simulations and compare it with the MSE approximations for the static (in [1]) and time-varying scenario (15).

2.3 Performance in the full dynamic scenario

So far, we have proved that introducing variance in the output process slows the attacker, increasing the MSE of the estimator, while variance in the input process leaves the users more vulnerable to the



(a) $t = 10$ (b) $t = 100$

Figure 2: MSE_p evolution with the number of rounds ρ, in an scenario where the sending frequencies vary in each round ($N = 100$, $n_f = 25$, $f_i = 1/N$, $t = 10$, 100).

least-squares disclosure attack. Intuitively, in the case where both effects coexist, we can expect the MSE curve to lay between the lower bound provided by the MSE with only time-varying frequencies (15) and the upper bound given by the MSE when only the profiles change between rounds (10). Therefore, as the number of observed rounds increases, the error would asymptotically reach zero.

We can prove this easily by showing that the variance of the estimator given the input observations decreases with ρ. First of all, we set

$$\begin{aligned} \boldsymbol{\Sigma}_{\hat{\mathbf{p}}_j|\mathbf{U}} &= (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\boldsymbol{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1} \\ &= \left(\frac{1}{\rho}\mathbf{U}^T\mathbf{U}\right)^{-1}\left(\frac{1}{\rho^2}\mathbf{U}^T\boldsymbol{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}\right)\left(\frac{1}{\rho}\mathbf{U}^T\mathbf{U}\right)^{-1} \end{aligned} \tag{16}$$

We can make the following statements:

- As long as the input process is ergodic, $\lim_{\rho\to\infty} \mathbf{U}^T\mathbf{U}/\rho \to \mathbf{R}_x$. Therefore, matrices $(\mathbf{U}^T\mathbf{U}/\rho)^{-1}$ will be approximately independent of ρ.
- On the other hand, the m,n-th element of $\mathbf{U}^T\boldsymbol{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}/\rho^2$ is

$$(\mathbf{U}^T\boldsymbol{\Sigma}_{\mathbf{Y}_j|\mathbf{U}}\mathbf{U}/\rho^2)_{m,n} = \frac{1}{\rho^2}\sum_{r=1}^{\rho}\sum_{s=1}^{\rho} X_m^r X_n^s \text{Cov}\left\{Y_j^r, Y_j^s|\mathbf{U}\right\} \tag{17}$$

  This term will decrease with ρ if the correlation time of the output process is finite, i.e., $\lim_{s\to\infty} \text{Cov}\left\{Y_j^r, Y_j^{r+s}\right\} \to 0$, or if this correlation decreases as $1/\rho$ or faster.

This means that, as long as the correlation time of the output process is finite or decreases fast, the variance of our estimator will decrease with ρ. The speed with which the MSE decreases will depend on how correlated the outputs are. This result is intuitive: when the outputs in different rounds are correlated, new observations provide less information to the attacker and therefore the estimation of the profiles becomes slower. An example of this would be the pool mix scenario.

References

[1] F. Pérez-González, C. Troncoso, and S. Oya, "A least squares approach to the static traffic analysis of high-latency anonymous communication systems," (Under submission).

[2] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 124–127, 1950.