

observable at the  $M$ th sampling instant, obviously it must be true that  $L \leq M - k$ . If more than  $Q$  temporally contiguous observables are available to the  $Q$ -tap Su-Wong-Ho estimator, guidelines are offered below to pick which  $Q$  contiguous observables for the estimator. This Section will rigorously prove that for a preset  $Q$ , the  $Q$ -tap data-window should centre around the  $k$ th time-sample as much as possible.

**Theorem 4.1:** Suppose that  $\{e^{j\theta_k}, k \leq M\}$  is to be estimated using a contiguous  $Q$ -tap observation-window taken from  $\{r_m, m = 1, 2, \dots, M\}$ . The following will minimise the MSE:

- (i) If  $k \in \{Q - 1/2, \dots, M - Q - 1/2\}$ , set  $L = \lfloor Q - 1/2 \rfloor$ .
- (ii) If  $k > M - Q - 1/2$ , set  $L = M - k$ .
- (iii) If  $k < Q - 1/2$ , set  $L = Q - k$  and  $e^{j\theta_k} = 1/A \sum_{q=1}^Q w_q r_q$

*Proof:* Define

$$\text{MMSE}(L) = \inf_{\mathbf{w}} \text{MSE}(L, \mathbf{w}) = \text{MSE}(L, \mathbf{w}^{(L,o)}) \quad (6)$$

where  $\mathbf{w}^{(L,o)} = [w_1^{(L,o)}, w_2^{(L,o)}, \dots, w_Q^{(L,o)}]$  denotes the LMMSE weights at latency  $L$ . Let  $L^*$  denote the 'best' latency over  $L = 0, 1, \dots, Q - 1$ , in the sense of

$$\text{MMSE}(L^*) = \min_L \text{MMSE}(L) \quad (7)$$

As shown in [3], the LMMSE weights  $w_1^{(L,o)}, w_2^{(L,o)}, \dots, w_Q^{(L,o)}$  must be positive real numbers, so the following will consider only  $w_q \in (0, \infty), \forall q = 1, \dots, Q$ . These LMMSE-estimator weights satisfy  $\partial/\partial w_q \text{MSE}(L, \mathbf{w}) = 0, \forall q$ :

$$w_q^{(L,o)} \left( \frac{\sigma_n^2}{A^2} \right) + \sum_{m=1}^Q w_m^{(L,o)} c^{|q-m|} = c^{|Q-L-q|}, \quad \forall q = 1, \dots, Q \quad (8)$$

Equivalently,

$$\sum_{m=1}^Q w_m^{(L,o)} c^{|q-m|} = c^{|Q-L-q|} - w_q^{(L,o)} \left( \frac{\sigma_n^2}{A^2} \right), \quad \forall q = 1, \dots, Q \quad (9)$$

$$\sum_{m=1}^Q w_m^{(L+1,o)} c^{|q-m|} = c^{|Q-L-1-q|} - w_q^{(L+1,o)} \frac{\sigma_n^2}{A^2}, \quad \forall q = 1, \dots, Q \quad (10)$$

As (5) implies

$$\text{MMSE}(Q - L - 1) = \text{MMSE}(L), \quad \forall L = 0, 1, \dots, Q - 1 \quad (11)$$

it would suffice to consider only  $L \leq \lfloor Q - 1/2 \rfloor$  below.

Suppose  $L < \lfloor Q - 1/2 \rfloor$ , or equivalently  $Q \geq 2L + 3$ . Then,

$$\begin{aligned} & \text{MMSE}(L) - \text{MMSE}(L + 1) \\ &= \text{MSE}(L, \mathbf{w}^{(L,o)}) - \text{MSE}(L + 1, \mathbf{w}^{(L+1,o)}) \\ &= \sum_{q=1}^Q (w_q^{(L,o)})^2 \left( \frac{\sigma_n^2}{A^2} \right) - 2 \left( 1 - w_{Q-L}^{(L,o)} \frac{\sigma_n^2}{A^2} \right) \\ & \quad + \sum_{q=1}^Q w_q^{(L,o)} \left( c^{|Q-L-q|} - w_q^{(L,o)} \frac{\sigma_n^2}{A^2} \right) \\ & - \left\{ \sum_{q=1}^Q (w_q^{(L+1,o)})^2 \left( \frac{\sigma_n^2}{A^2} \right) - 2 \left( 1 - w_{Q-L-1}^{(L+1,o)} \frac{\sigma_n^2}{A^2} \right) \right. \\ & \quad \left. + \sum_{q=1}^Q w_q^{(L+1,o)} \left( c^{|Q-L-q|} - w_q^{(L+1,o)} \frac{\sigma_n^2}{A^2} \right) \right\} \\ &= \left( w_{Q-L}^{(L,o)} - w_{Q-L-1}^{(L+1,o)} \right) \frac{\sigma_n^2}{A^2} \\ &= \left[ \frac{(\alpha_{L-1} - \alpha_L - (\alpha_{Q-L-2} - \alpha_{Q-L-1}))}{(\beta - \alpha_{Q-L-1} - \alpha_L)(\beta - \alpha_{Q-L-2} - \alpha_{L+1})} \right] \frac{1 - c^2}{c} \\ &> 0 \end{aligned} \quad (12)$$

The above (12) is obtained by setting  $q = Q - L$  in (9) and setting  $q = Q - L - 1$  in (10). That gives

$$\begin{aligned} \sum_{q=1}^Q w_q^{(L,o)} c^{|Q-L-q|} &= 1 - w_{Q-L}^{(L,o)} \frac{\sigma_n^2}{A^2} \\ \sum_{q=1}^Q w_q^{(L+1,o)} c^{|Q-L-1-q|} &= 1 - w_{Q-L-1}^{(L+1,o)} \frac{\sigma_n^2}{A^2} \end{aligned}$$

To prove (13), it would be sufficient to show that  $(\alpha_{L-1} - \alpha_L) > (\alpha_{Q-L-2} - \alpha_{Q-L-1})$ . Towards this end, note that  $\alpha_{q+1} = 1/\beta - \alpha_q \forall q \geq 0$  is equivalent to  $\beta - \alpha_q = 1/\alpha_{q+1}$  and  $\alpha_q = \beta - 1/\alpha_{q+1}, \forall q \geq 0$ . So,  $(\alpha_{L-1} - \alpha_L) - (\alpha_{Q-L-2} - \alpha_{Q-L-1}) = (\beta - 1/\alpha_L - \alpha_L) - (\beta - 1/\alpha_{Q-L-1} - \alpha_{Q-L-1}) = \alpha_L - \alpha_{Q-L} - 1/\alpha_L \alpha_{Q-L-1} (1 - \alpha_L \alpha_{Q-L-1})$ . Because  $Q \geq 2L + 3$  and  $\alpha_m$  both strictly decrease with  $m$ , it holds that  $Q - L - 1 > L$  and  $1 \geq \alpha_L > \alpha_{Q-L-1} > 0$ . This proves (13) and completes the proof of the theorem.  $\square$

*Recursive update of Su-Wong-Ho estimator weights to incorporate new sample:* For a  $Q$ -tap estimator, what happens if both the  $Q$ -tap observation-window and the to-be-estimated time-sample ( $e^{j\theta_k}$ ) slide to the right along the discrete-time axis, but  $L$  remains the same? How may the correspondingly right-shifted Su-Wong-Ho LMMSE estimator's weights be related to the original 'pre-shift' Su-Wong-Ho LMMSE estimator's weights? The LMMSE estimator's  $Q$ -taps  $\{w_q^{(o)}, q = 1, \dots, Q\}$ , that respectively weighting the observables  $\{r_{k+L-Q+q}, q = 1, \dots, Q\}$  in  $e^{j\theta_k}$ , those taps would be applicable also in  $e^{j\theta_{k+1}}$  to weight  $\{r_{k+L-Q+q+1}, q = 1, \dots, Q\}$  after the aforementioned 'shifting'. This is because

- (i) between  $\theta_{k+1}$  and  $\theta_{M+1}$  is a time lapse equal to that between  $\theta_k$  and  $\theta_M$ , and
- (ii) the phase noise is modelled to have stationary independent increments.

The weights are thus independent of  $k$ , even though  $\{\theta_k\}$  and  $\{e^{j\theta_k}\}$  are non-stationary. This independence of  $k$  is because the estimator multiplies  $r_k^*$  with  $r_m, \forall m$ .

*Conclusion:* The Su-Wong-Ho phase-estimator of [3] was for a limited range of latency, and for a pre-specified number of taps equal to the observation-window size. That estimator is herein extended to accommodate a wider range of latency, to select the optimum estimator-taps window from a wider observation-window, and to handle newly arrived samples.

*Acknowledgment:* This work was supported by the Internal Competitive Research Grant number G-YG67 from the Hong Kong Polytechnic University.

© The Institution of Engineering and Technology 2011

1 March 2011

doi: 10.1049/el.2011.0541

Y.-T. Su (*Department of Mathematics, National Kaohsiung Normal University, Kaohsiung 80264, Taiwan*)

E-mail: yts@nknuc.nknu.edu.tw

K.T. Wong (*Department of Electronic & Information Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*)

## Reference

- 1 Su, Y.-T., Wong, K.T., and Ho, K.-P.: 'Linear MMSE estimation of large-magnitude symmetric Levy-process phase-noise', *IEEE Trans. Commun.*, 2010, **58**, (12), pp. 3369-3375

## Average Fisher information maximisation in presence of cost-constrained measurements

B. Dulek and S. Gezici

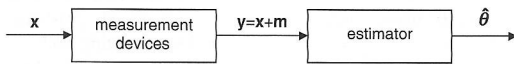
An optimal estimation framework is considered in the presence of cost-constrained measurements. The aim is to maximise the average Fisher information under a constraint on the total cost of measurement devices. An optimisation problem is formulated to calculate the optimal costs of measurement devices that maximise the average Fisher information for arbitrary observation and measurement statistics. In addition, a closed-form expression is obtained in the case of Gaussian observations and measurement noise. Numerical examples are presented to explain the results.

*Introduction:* In estimation problems, the Cramer-Rao lower bound (CRLB) provides a lower bound on mean-squared errors (MSEs) of

unbiased estimators. In addition, when the prior distribution of the unknown parameter is known, the Bayesian CRLB (BCRLB) can be calculated to obtain a lower bound on the MSE of any estimator [1]. The CRLB and BCRLB are quite useful in the analysis of estimation problems since (a) they provide lower bounds that can (asymptotically) be achieved by certain estimators (e.g. the maximum likelihood estimator), (b) they are easier to calculate than the MSE as their formulations do not depend on any specific estimator structure. Recently, a novel measurement device model has been proposed, and the problem of designing the optimal linear estimator is studied under a total cost constraint on the measurement devices [2]. Unlike previous studies, it is considered that each observation is measured by a measurement device, the accuracy of which depends on the cost spent on that device. In that way, a total cost constraint is taken into account and the optimal linear estimator design is performed under that constraint.

In this Letter, we consider the problem of minimising the BCRLB (equivalently, maximising the average Fisher information) at the outputs of measurement devices under the total cost constraint introduced in [2]. In other words, we propose a generic formulation for determining the optimal cost allocation among measurement devices in order to maximise the average Fisher information. We also obtain a closed-form expression for the Gaussian case, and present numerical examples.

**Optimal solution:** Consider a scenario as in Fig. 1 in which a  $K$ -dimensional observation vector  $\mathbf{x}$  is measured by  $K$  measurement devices, and then the measured values in vector  $\mathbf{y}$  are processed to estimate the value of parameter  $\theta$ . The measurement devices are modelled to introduce additive measurement noise denoted by  $\mathbf{m}$ . In other words, the probability density function (PDF) of  $\mathbf{x}$  is indexed by parameter  $\theta$ , and the aim is to estimate that parameter based on the outputs of the measurement devices. Although a linear system model and a different problem formulation are considered in [2], motivations for that study can also be invoked for the system model in Fig. 1. It should be emphasised that the model in Fig. 1 presents a generic estimation framework in which measurements are processed by an estimator in order to determine the value of an unknown parameter. For example, in a wireless sensor network application, measurement devices correspond to sensors, which are used to estimate a parameter in the system, such as the temperature.



**Fig. 1** Observation vector  $\mathbf{x}$  measured by  $K$  measurement devices, and measurements  $\mathbf{x} + \mathbf{m}$  are used by estimator to estimate value of unknown parameter  $\theta$

To consider practical system constraints, we assume that there is a total cost constraint on the measurement devices, as proposed in [2]. Specifically, the total cost budget of the measurement devices cannot exceed  $C$ , which is specified by

$$\frac{1}{2} \sum_{i=1}^K \log \left( 1 + \frac{\sigma_{x_i}^2}{\sigma_{m_i}^2} \right) \leq C \quad (1)$$

where  $\sigma_{x_i}^2$  denotes the variance of the  $i$ th component of observation vector  $\mathbf{x}$ , and  $\sigma_{m_i}^2$  is the variance of the  $i$ th measurement device (i.e., the  $i$ th component of  $\mathbf{m}$ ). In other words, it is assumed that a measurement device has a higher cost if it can perform measurements with a lower measurement variance (i.e. with higher accuracy). Various motivations for the cost constraint in (1) can be found in [2].

To maximise the estimation accuracy, we consider the maximisation of average Fisher information, or equivalently the minimisation of the BCRLB at the output of the measurement devices. The main motivation for the suggested approach is that an optimal cost assignment strategy can be obtained by solving such an optimisation problem without assuming a specific estimator structure. In addition, it is known that some estimators, such as the maximum *a-posteriori* probability estimator, can (asymptotically) achieve the BCRLB; hence, the minimisation of the BCRLB corresponds to the (approximate) minimisation of the MSE for certain estimators.

For an arbitrary estimator  $\hat{\theta}$ , the BCRLB on the MSE is expressed as [1]

$$\text{MSE}\{\hat{\theta}\} = \text{E}\{(\hat{\theta}(\mathbf{y}) - \theta)^2\} \geq (\mathbf{J}_D + \mathbf{J}_P)^{-1} \quad (2)$$

where  $\mathbf{J}_D$  and  $\mathbf{J}_P$  denote the information obtained from observations and prior knowledge, respectively, which are stated as

$$\mathbf{J}_D = \text{E} \left\{ \left( \frac{\partial \log p_{\mathbf{Y}}^{\theta}(\mathbf{y})}{\partial \theta} \right)^2 \right\}, \quad \mathbf{J}_P = \text{E} \left\{ \left( \frac{\partial \log w(\theta)}{\partial \theta} \right)^2 \right\} \quad (3)$$

with  $p_{\mathbf{Y}}^{\theta}(\mathbf{y})$  and  $w(\theta)$  representing the PDF of  $\mathbf{Y}$  and the prior PDF of the parameter, respectively. As  $\mathbf{J}_P$  depends only on the prior PDF, it is independent of the cost of the measurement devices. Therefore, the aim is to maximise  $\mathbf{J}_D$ , which is defined as the average Fisher information, under the cost constraint in (1). To specify this optimisation problem, it is assumed that the observation is independent of the measurement noise; hence,  $p_{\mathbf{Y}}^{\theta}(\mathbf{y})$  in (3) can be expressed more explicitly as the convolution of the PDFs of  $\mathbf{x}$  and  $\mathbf{m}$ ; i.e.  $p_{\mathbf{Y}}^{\theta}(\mathbf{y}) = \int p_{\mathbf{X}}^{\theta}(\mathbf{y} - \mathbf{m}) p_{\mathbf{M}}(\mathbf{m}) d\mathbf{m}$ . In addition, it is reasonable to assume that each measurement device introduces independent noise, in which case  $p_{\mathbf{M}}(\mathbf{m})$  becomes  $p_{\mathbf{M}}(\mathbf{m}) = p_{M_1}(m_1) \dots p_{M_K}(m_K)$ . As discussed in [2], the cost of a measurement device can be expressed as a function of its measurement noise variance (see (1)). Each measurement noise component can be modelled as  $m_i = \sigma_{m_i} \tilde{m}_i$ , where  $\tilde{m}_i$  denotes a zero-mean unit-variance random variable with a known PDF  $p_{\tilde{M}_i}$ , and  $\sigma_{m_i}^2$  represents the variance of the measurement device, which determines its cost as defined in (1). Hence, the PDF of the  $i$ th measurement noise can be expressed as  $p_{M_i}(m) = \sigma_{m_i}^{-1} p_{\tilde{M}_i}(\sigma_{m_i}^{-1} m)$ .

Based on (1) and (3), the optimal cost assignment problem can be formulated as

$$\begin{aligned} & \max_{\{\sigma_{m_i}^2\}_{i=1}^K} \int_{\mathcal{R}^K} w(\theta) \int_{\mathcal{R}^K} p_{\mathbf{Y}}^{\theta}(\mathbf{y}) \left( \frac{dp_{\mathbf{Y}}^{\theta}(\mathbf{y})}{d\theta} \right)^2 d\mathbf{y} \\ & \text{subject to } \frac{1}{2} \sum_{i=1}^K \log \left( 1 + \frac{\sigma_{x_i}^2}{\sigma_{m_i}^2} \right) \leq C \end{aligned} \quad (4)$$

It is noted that the expectation operator for the calculation of  $\mathbf{J}_D$  in (3) is over both  $\theta$  and  $\mathbf{Y}$ , resulting in the objective function in (4). From the discussions in the previous paragraph, we have  $p_{\mathbf{Y}}^{\theta}(\mathbf{y}) = \int p_{\mathbf{X}}^{\theta}(\mathbf{y} - \mathbf{m}) \prod_{i=1}^K \sigma_{m_i}^{-1} p_{\tilde{M}_i}(\sigma_{m_i}^{-1} m_i) d\mathbf{m}$ , which becomes

$$\begin{aligned} p_{\mathbf{Y}}^{\theta}(\mathbf{y}) &= \prod_{i=1}^K \sigma_{m_i}^{-1} \int_{-\infty}^{\infty} p_{X_i}^{\theta}(y_i - m_i) p_{\tilde{M}_i}(\sigma_{m_i}^{-1} m_i) dm_i \\ &= \prod_{i=1}^K \int_{-\infty}^{\infty} p_{X_i}^{\theta}(y_i - \sigma_{m_i} m) p_{\tilde{M}_i}(m) dm \end{aligned} \quad (5)$$

in the case of independent observations. In fact, the objective function in (4) can be written as the sum of  $K$  components in that case (see (3)) as

$$\sum_{i=1}^K \int w(\theta) \int_{-\infty}^{\infty} \frac{1}{p_{Y_i}^{\theta}(y)} \left( \frac{dp_{Y_i}^{\theta}(y)}{d\theta} \right)^2 dy$$

where

$$p_{Y_i}^{\theta}(y) = \int_{-\infty}^{\infty} p_{X_i}^{\theta}(y - \sigma_{m_i} m) p_{\tilde{M}_i}(m) dm$$

Since the optimisation problem in (4) provides a generic formulation that is valid for any observation PDF, the problem can be non-concave in general. Hence, global optimisation tools such as particle swarm optimisation and differential evolution can be used to obtain the solution [3].

**Special case:** In the case of independent Gaussian observations and measurement noise, it is possible to obtain closed-form solutions of the optimisation problem in (4). To that aim, let the observation vector  $\mathbf{x}$  have independent Gaussian components denoted as  $X_i \sim N(\theta, \sigma_{x_i}^2)$  for  $i = 1, \dots, K$ , and let each measurement noise component have independent zero-mean Gaussian distribution with variance  $\sigma_{m_i}^2$ . In that case, the average Fisher information  $\mathbf{J}_D$  can be calculated as  $\sum_{i=1}^K (\sigma_{m_i}^2 + \sigma_{x_i}^2)^{-1}$ . Hence, the aim becomes the maximisation of  $\sum_{i=1}^K (\sigma_{m_i}^2 + \sigma_{x_i}^2)^{-1}$  over  $\sigma_{m_1}^2, \dots, \sigma_{m_K}^2$  under the constraint in (1). It is noted that both the objective function and the constraint are convex in this optimisation problem. Since the maximum of convex functions over convex sets has to occur at the boundary [4], the cost constraint becomes equality, and the solution of the optimisation problem can be obtained by using Lagrange multipliers [4], resulting in the following

algorithm for the optimal cost allocation:

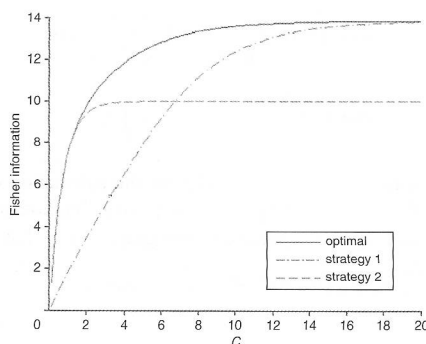
$$\sigma_{m_i}^2 = \begin{cases} \frac{\sigma_{x_i}^A}{\gamma - \sigma_{x_i}^2}, & \text{if } \sigma_{x_i}^2 < \gamma \\ \infty & \text{if } \sigma_{x_i}^2 \geq \gamma \end{cases} \text{ with } \gamma = \left( 2^{2C} \prod_{i \in S_K} \sigma_{x_i}^2 \right)^{1/|S_K|} \quad (6)$$

where  $S_K = \{i \in \{1, \dots, K\} : \sigma_{m_i}^2 \neq \infty\}$  and  $|S_K|$  denotes the number of elements in set  $S_K$ . In other words, if the observation noise variance is larger than a threshold  $\gamma$ , a measurement device with infinite variance (that is, with zero cost) is considered; namely, that observation is not measured at all. On the other hand, for observations with variances smaller than  $\gamma$ , the noise variance of the corresponding measurement device is determined according to the formulation in (6), which assigns low measurement variances (high costs) to observations with low variances.

**Table 1:** Measurement variances and corresponding Fisher information for optimal strategy (see (6)), strategy 1, strategy 2

	$\sigma_{m_1}^2$	$\sigma_{m_2}^2$	$\sigma_{m_3}^2$	$\sigma_{m_4}^2$	Fisher information
Optimal	0.0097	0.3973	3.533	$\infty$	10.45
Strategy 1	0.4373	0.4373	0.4373	0.4373	4.252
Strategy 2	0.0032	$\infty$	$\infty$	$\infty$	9.688

Parameters are  $C = 2.5$ ,  $\sigma_{x_1}^2 = 0.1$ ,  $\sigma_{x_2}^2 = 0.5$ ,  $\sigma_{x_3}^2 = 0.9$ , and  $\sigma_{x_4}^2 = 1.3$



**Fig. 2** Fisher information against total cost  $C$  for optimal strategy, strategy 1, strategy 2 where  $\sigma_{x_1}^2 = 0.1$ ,  $\sigma_{x_2}^2 = 0.5$ ,  $\sigma_{x_3}^2 = 0.9$ , and  $\sigma_{x_4}^2 = 1.3$

— optimal  
- - - strategy 1  
- . - . strategy 2

**Alternative strategies:** Instead of the optimal cost assignment strategy specified in (4), one can also consider the following simple alternatives.

**Strategy 1** (equal measurement device variances): In this strategy, it is assumed that measurement devices with equal variances are used for all observations; i.e.  $\sigma_{m_i}^2 = \sigma_m^2$ ,  $i = 1, \dots, K$ . Then, the cost constraint in (4) can be used with equality, and  $\sigma_m^2$  is simply obtained as the smallest positive real root of  $\prod_{i=1}^K (1 + \sigma_{x_i}^2 / \sigma_m^2) = 2^{2C}$ . If the observation variances are also equal,  $\sigma_m^2$  becomes  $\sigma_m^2 = \sigma_x^2 / (2^{2C/K} - 1)$ .

**Strategy 2** (all cost to the best observation): In this case, the total budget  $C$  is spent on the best observation, which has the smallest variance. If the  $b$ th observation is the best one, the cost constraint in (4) can be used to calculate the variance of the measurement noise for that observation as  $\sigma_{m_b}^2 = \sigma_{x_b}^2 / (2^{2C} - 1)$ . For all the other observations, the corresponding measurement variances are set to infinity (i.e. no measurements are taken from those observations).

**Results and conclusions:** To provide numerical examples of the results in the preceding Sections, consider a scenario with independent Gaussian observations and measurement noise. Let  $C = 2.5$ ,  $\sigma_{x_1}^2 = 0.1$ ,  $\sigma_{x_2}^2 = 0.5$ ,  $\sigma_{x_3}^2 = 0.9$ , and  $\sigma_{x_4}^2 = 1.3$ . In Table 1, the variances of the measurement devices and the corresponding Fisher information values are shown for the proposed optimal strategy (see (6)), strategy 1 and strategy 2. It is observed that the optimal strategy assigns smaller variances (larger costs) to observations with smaller variances, and achieves the maximum Fisher information as expected. For further investigations, Fig. 2 illustrates the Fisher information versus the total budget  $C$  for different strategies. It is observed that the Fisher

information in strategy 2, which assigns all the cost to the best observation, converges to  $1/\sigma_{x_1}^2$  as expected (since  $\sigma_{m_1}^2$  converges to zero as  $C$  increases). On the other hand, strategy 2 and the optimal strategy converge for very small values of  $C$  since the optimal strategy involves assigning all the cost to the best observation if  $C$  is small. Regarding strategy 1, it converges to the optimal strategy for large  $C$ , and significant deviations are observed for intermediate values of  $C$ . Overall, the optimal cost assignment strategy yields the highest Fisher information in all the cases, and indicates the opportunity to achieve high estimation accuracy.

© The Institution of Engineering and Technology 2011

11 March 2011

doi: 10.1049/el.2011.0686

One or more of the Figures in this Letter are available in colour online.

B. Dulek and S. Gezici (*Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara 06800, Turkey*)

E-mail: gezici@ee.bilkent.edu.tr

## References

- 1 Van Trees, H.L.: 'Detection, estimation, and modulation theory, Part I' (Wiley-Interscience, New York, 2001)
- 2 Ozcelikkale, A., Ozaktas, H.M., and Arikan, E.: 'Signal recovery with cost-constrained measurements', *IEEE Trans. Signal Process.*, 2010, **58**, (7), pp. 3607–3617
- 3 Storn, R., and Price, K.: 'Differential evolution – a simple and efficient heuristic for global optimisation over continuous spaces', *J. Global Optimiz.*, 1997, **11**, pp. 341–359
- 4 Rockafellar, R.: 'Convex analysis' (Princeton University Press, 1972)

## Detecting DNS-poisoning-based phishing attacks from their network performance characteristics

H. Kim and J.H. Huh

Most of the existing phishing detection techniques are weak against domain name system (DNS)-poisoning-based phishing attacks. Proposed is a highly effective method for detecting such attacks: the network performance characteristics of websites are used for classification. To demonstrate how useful the approach is, the performance of four classification algorithms are explored: linear discriminant analysis, naïve Bayesian,  $K$ -nearest neighbour, and support vector machine. Over 10 000 real-world items of routing information have been observed during a one-week period. The experimental results show that the best-performing classification method – which uses the  $K$ -nearest neighbour algorithm – is capable of achieving a true positive rate of 99.4% and a false positive rate of 0.7%.

**Introduction:** Over the years, many techniques have been developed to help identify phishing websites; some of these include whitelisting of legitimate sites, blacklisting of fraudulent sites and identifying common heuristics. Unfortunately, when it comes to detecting domain name system (DNS)-poisoning-based phishing attacks, such techniques tend to achieve low performance [1]: Abu-Nimeh and Nair introduced an attacking scenario in which a rogue access point (with stronger signal range) is setup at Starbucks, and a user connects to the Internet through this access point rather than Starbucks' hotspot [1]. As the local DNS in this access point is poisoned, when the user types a URL in her browser, she is directed to a phishing website hosted at the access point's local server. Here, her usual security toolbars and phishing filters will not provide any warning message. This is because most of these techniques merely check the domain names of websites and/or rely on remote verification servers to perform the necessary phishing detection tasks. In practice, however, the communication channel between client software and a remote server is not often secured, allowing attackers to easily forge the server's response. Keeping the IP addresses alongside the domain names could be effective only if the IP addresses were all static. Content-based phishing detection methods too have inherent weaknesses: phishing contents can be sophisticatedly created to avoid the (publicly known) properties that are being checked by heuristics.