Progress In Electromagnetics Research, Vol. 136, 457–478, 2013

RECONSTRUCTION OF FAULTY CABLE NETWORK USING TIME-DOMAIN REFLECTOMETRY

Xiaolong Zhang, Minming Zhang*, and Deming Liu

National Engineering Laboratory for Next Generation Internet Access System, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract—Based on Time-Domain Reflectometry (TDR) technique, a novel method which could locate faults on the coaxial cable distribution network by using Support Vector Machine (SVM) is proposed in this paper. This approach allows the faulty network to be reconstructed by estimating the lengths of branches. A State-transition Matrix model is employed to simulate the TDR response at any port and evaluate the transfer function between two points. SVM is used to solve the inversion problem through training datasets created by the State-transition matrix model. Compared to the existing reflectometry methods, our proposed method can tackle multiple faults in the complex cable networks. Numerical and experimental results pointing out the performance of the SVM model in locating faults are reported.

1. INTRODUCTION

The coaxial cable network is widely used in the multi-service access for residential subscribers, such as CATV (Community Antenna Television), HFC (Hybrid Fiber Coaxial), DOCSIS (Data Over Cable System Interface Specification), and etc.. Moreover, the process of standardization of EPoC (EPON over Coax) in the IEEE 802.3 Working Group is currently underway. At the time of writing, a Study Group has been formed to examine EPoC as well as the cable network [1, 2]. In order to guarantee the quality of service to users, the diagnosis and location of fault on the network is a crucial common problem.

Generally, there are two main kinds of cable network faults: "hard faults" and "soft faults". The hard faults are mostly open or short

Received 14 December 2012, Accepted 10 January 2013, Scheduled 21 January 2013
* Corresponding author: Minming Zhang (kindmercy@163.com).

circuits, which lead to a strong change when the signal propagates through the network. The soft faults are caused by discontinuities of the impedance in network, e.g., damaged insulation and water infiltration. If the soft faults are not repaired for a long time, they can eventually develop into hard faults. Therefore, it is important to detect and locate faults as early as possible.

During the last decade, many techniques have been proposed to locate defects in cables [3–5]. Among them, reflectometry method is the most important one and still widely used today. For instance, a number of declinations of reflectometry method have been proposed to improve the location capacities by using optimized testing signals and developing more accurate reflectometry method [6]. Although these methods have better capabilities to locate fault on single cable, it is hard for them to locate faults on the branched network. In practice, the coaxial cables are always used in distribution network with other components such as amplifier, splitter, tap, and etc.. Hence, a new method which could locate faults on real cable network is highly desirable.

Inversion problem is another issue for the reflectometry method. The junctions and ends of branched network all result in multiple reflections in the reflectometry trace, so it is difficult to extract fault location. Hence, it requires an intelligent algorithm to locate the fault on branched network from the reflectometry trace. With the development of artificial intelligence technology, artificial neural network (ANN) with strong nonlinear mapping and robust ability has been widely applied to solve the inversion problem and locate fault [7, 8]. However, there is no universal method to determine an optimal ANN structure in terms of the number of hidden layers and number of neurons in each layer. Moreover, ANN has the shortcomings of overfitting and sinking into the local optimal. Compared to traditional ANN, SVM (Support Vector Machine) exhibits the major advantage of global optimization, higher generalization capability and input-dimension independence. Alternatively, SVM unfold a promising means to estimate nonlinear system models accurately [9–14].

In this paper, an efficient method, based on TDR (Time-Domain Reflectometry) and SVM, is presented to locate multiple faults on complex cable network. A State-transition Matrix model is introduced to simulate the TDR response at any port and evaluate the transfer function between two points. The inversion problem is solved by the SVM. Comparing with other common methods which require accurate length of each cable or just locate fault on single cable, this method allows the faulty network to be reconstructed by estimating the lengths of branches, only need to know the topology of distribution network.

The aim of the paper is finding the relationship between reflectometry response and topology network, and solving the inversion problem by SVM to reconstruct the network. Since this method is operated with topology-described matrixes, it is easy to implement the process by computer and develop it into embedded system.

The paper is organized as follows. First, the numerical model of TDR and the inversion model by SVM are built in Section 2, including the basic ideas and operating procedure. Next, the algorithm's effectiveness is demonstrated and the estimated and measured results obtained from actual coaxial cable distribution networks are presented in Section 3. Finally, Section 4 gives the conclusion.

2. THEORY OF THE PROPOSED METHOD

The proposed method can be used to locate multiple faults on complex cable network. First, the topology of analyzed faulty network need to ascertain (without accurate lengths of cables), and each model for component in the network is built, including the coaxial cable, splitter, tap and so on. Next, State-transition Matrix is introduced to estimate the TDR responses of networks with different fault locations. After that, the TDR curves can be used to train the SVM. Finally, once being correctly trained, the SVM can locate faults from the measured TDR trace.

2.1. Numerical Model for Each Component

When the topology of network is definite, the components are also determined. In most residential areas, the topology of the cable distribution network is the same, the difference is only the length of cable. Generally, a coaxial cable distribution network consists of many coaxial cables and some devices, such as splitters, taps, and various loads. According to the transmission line theory, the propagation constant and transfer function of a cable with length l can be simply expressed by

$$\gamma = k_1\sqrt{f} + k_2 f + j \cdot 2\pi f/v \tag{1}$$

$$H(f,l) = e^{-\gamma \cdot l} = e^{-(k_1\sqrt{f}+k_2 f+j\cdot 2\pi f/v)\cdot l} \tag{2}$$

where f is the frequency, v the velocity of propagation, and k_1, k_2 are constants and depended on the type of cable [15].

Based on the measurement of a certain length cable by the Vector Network Analysis (VNA), the amplitude-frequency curve and phase-frequency curve are obtained. Using a least-squares fitting algorithm, parameters k_1, k_2 and v can be determined through (1). In this paper,

the same type of cable (RG-6) are used to construct the network, and their regressive parameters are $k_1 = 8.53e-6$, $k_2 = -2.95e-10$, $v = 2.45e+8$, and the characteristic impedance Z_0 is 75 Ω.

Other important devices in the cable network are splitters and taps, etc.. Generally, there are two different models for these devices: the first one is theoretically calculated model; the second one is the database model. For greatest accuracy, the second one is chosen and the actual scattering matrixes of devices are measured by VNA. In this paper, the frequency range is from 1 MHz to 1.2 GHz, so that the scope of measurement should cover this frequency range. Moreover, the measured scattering matrix should be changed to the reference characteristic impedance of cable, i.e., Z_0.

2.2. State-transition Matrix for Complex Network Topology

When a signal is sent down the cable network, it will propagate through numerous paths from transmitter to receiver. Those multiple paths are generated by the impedance mismatch terminals, ports of devices and faults. Each path is characterized by a propagation loss factor $e^{-\alpha l_i}$, weighting g_i, and a delay t_i, the transfer function from transmitter to receiver can be expressed as follows [16]:

$$H(f) = \lim_{n\to\infty} \sum_{i=1}^{N} g_i \cdot e^{-\alpha l_i} \cdot e^{-j2\pi f t_i} = \lim_{n\to\infty} \sum_{i=1}^{N} g_i \cdot e^{-\gamma l_i} \tag{3}$$

where l_i is the length of the path i, γ the propagation constant expressed in (1), and N the number of paths. The weighting factor g_i is a complex number which can be expressed as the product of all reflection and transmission coefficients, it can be expressed as follows:

$$g_i = \prod_{k=1}^{M_1} \rho_{ik} \prod_{n=1}^{M_2} \tau_{in} \tag{4}$$

where M_1 and M_2 are the number of reflection and transmission coefficients in the path i, respectively.

As mentioned above, the scattering matrix at each interested frequency has been changed to Z_0, so that the reflection coefficient ρ, at the port j of a device, is equal to the element S_{jj} in scattering matrix, and the transmission coefficient τ from the port i to j is equal to S_{ji}. Moreover, the reflection and transmission can be occurred at the mismatch terminal or fault. According to the basic transmission line theory, the reflection and transmission coefficient are given as

$$\rho = (Z_L - Z_0)/(Z_L + Z_0) \tag{5}$$

$$\tau = 1 + \rho = (2Z_L)/(Z_L + Z_0) \tag{6}$$

where Z_L is the impedance of the mismatch terminal or fault and Z_0 the characteristic impedance of cable.

In order to estimate the transfer function expressed in (3), a novel algorithm is developed to calculate the all paths. The algorithm treats the network as a finite state flow graph. A state is defined by two variables: one is the last cable that the signal has propagated along; another is the current node that the signal is in. The state is denoted as (*last cable — current node*). When the signal propagates from one state to another, a state transition occurs. So a signal will go through many state transitions from transmitter to receiver. For example, shown in Fig. 1 is a Y-type topology where three cables (RG-6) are connected together by a splitter ("203"). As the reflectometry, there are many paths from transmitter T_1 and back to receiver T_1, such as the path expressed as follows:

$$i = 1 : T_1\text{-}(L_1\text{-}C_1)\text{-}(L_2\text{-}T_2)\text{-}(L_2\text{-}C_1)\text{-}(L_1\text{-}T_1) \tag{7}$$

where each part in a bracket is a state, and the frequency response of this path can expressed by

$$H_1(f) = \left(S_{21}e^{-\gamma\cdot L_1}\right)\left(\rho_{T_2}e^{-\gamma\cdot L_2}\right)\left(S_{12}e^{-\gamma\cdot L_2}\right)\left(\tau_{T_1}e^{-\gamma\cdot L_1}\right) \tag{8}$$

where each expression in a bracket is the corresponding frequency response of a state transition, e.g., the element $S_{21}e^{-\gamma\cdot L_1}$, is happened when the signal propagates from state L_1C_1 to the state L_2T_2.

In order to calculate and cumulate the all paths expressed in (3), all the allowable states should be found out in the topology network. Supposing that there is a total number of N states in the topology network (in fact, N is twice of the numbers of cables), they are arranged as sequence $\mathbf{Q}$, the corresponding lengths of the cables in these states are arranged in a vector $\mathbf{L}$, i.e.,

$$\mathbf{Q} = [q_1 \quad q_2 \quad \dots \quad q_N] \tag{9}$$

$$\mathbf{L} = [l_1 \quad l_2 \quad \dots \quad l_N] \tag{10}$$

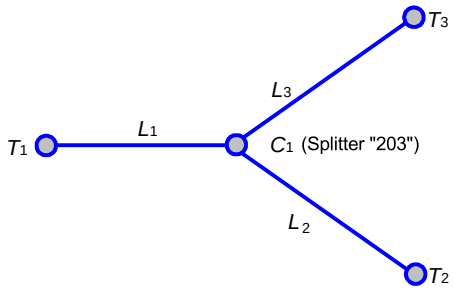

Figure 1. A simple topology of coaxial cable network.

If the frequency responses of the state transitions between any two states in the sequence $\mathbf{Q}$ can be calculated, a $N \times N$ state-transition matrix $\mathbf{G}$ can be formulated with each row and column corresponding to a state, i.e.,

$$\mathbf{G} = \begin{bmatrix} a_{11}e^{-\gamma \cdot l_1} & a_{12}e^{-\gamma \cdot l_2} & \dots & a_{1N}e^{-\gamma \cdot l_N} \\ a_{21}e^{-\gamma \cdot l_1} & a_{22}e^{-\gamma \cdot l_2} & \dots & a_{2N}e^{-\gamma \cdot l_N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}e^{-\gamma \cdot l_1} & a_{N2}e^{-\gamma \cdot l_2} & \dots & a_{NN}e^{-\gamma \cdot l_N} \end{bmatrix} \tag{11}$$

where the rows of the matrix are labeled by the current state, and the columns are labeled by the next state. Each element in $\mathbf{G}$ is frequency response of a state transition, i.e., $a_{ij}e^{-\gamma \cdot l_j}$ is the frequency response of a state transition from state q_j to state q_i. So $\mathbf{G}$ includes all the state transitions and indicates how the signal will propagate in the coaxial cable network.

2.3. Reflectometry Response of Complex Network Topology

It is very interesting that, the power of $\mathbf{G}$ has an important property: the element $\mathbf{G}^k(i,j)$ expresses the sum of frequency responses of all possible paths through k state transitions, and each path is from the beginning state q_j to the end state q_i. Supposing that q_j and q_i are corresponding to the transmitter T_t and receiver T_r, and the length of the cable connected to the receiver is l_r, the transmission coefficient at the receiving terminal is τ_r. So multiplying $\mathbf{G}^k(i,j)$ by $\tau_r e^{-\gamma \cdot l_r}$, the corresponding sum of responses with $k+1$ state transitions from T_t to T_r are obtained:

$$\begin{aligned} H'_{k+1}(i,j) &= \tau_r e^{-\gamma \cdot l_r} \cdot \mathbf{G}^k(i,j) \\ &= \sum_{n_{k-1}=1}^{N} \sum_{n_{k-2}=1}^{N} \cdots \sum_{n_2=1}^{N} \sum_{n_1=1}^{N} a_{in_{k-1}} a_{n_{k-1}n_{k-2}} \cdots a_{n_2 n_1} a_{n_1 j} \\ &\quad \times e^{-\gamma \cdot (l_i + l_{n_{k-1}} + l_{n_{k-2}} + \cdots + l_{n_2} + l_{n_1})} \tau_r e^{-\gamma \cdot l_r} \end{aligned} \tag{12}$$

It is supposed that direct connection between terminals does not exist. In this respect, the frequency response of all paths from transmitter T_t to receiver T_r can be expressed as:

$$\begin{aligned} H'(i,j) &= \sum_{k=2}^{\infty} H'_k(i,j) = \tau_r e^{-\gamma \cdot l_r} \cdot \sum_{k=1}^{\infty} \mathbf{G}^k(i,j) \\ &= \tau_r e^{-\gamma \cdot l_r} \cdot \left[\sum_{k=1}^{\infty} \mathbf{G}^k \right](i,j) = \tau_r e^{-\gamma \cdot l_r} \cdot \left[\mathbf{G} \cdot [\mathbf{E} - \mathbf{G}]^{-1} \right](i,j) \end{aligned} \tag{13}$$

where $\mathbf{E}$ is the identity matrix.

For convenience, an ending-state-transition matrix $\mathbf{D}$ is formed, i.e.,

$$\mathbf{D} = \text{Diag}\left[\begin{array}{cccc} \tau_1 e^{-\gamma \cdot l_1} & \tau_2 e^{-\gamma \cdot l_2} & \cdots & \tau_N e^{-\gamma \cdot l_N} \end{array}\right] \tag{14}$$

Pre-multiplying $\mathbf{G} \cdot [\mathbf{E} - \mathbf{G}]^{-1}$ by $\mathbf{D}$, an innovative matrix can be obtained

$$\mathbf{H} = \mathbf{D} \cdot \mathbf{G} \cdot [\mathbf{E} - \mathbf{G}]^{-1} \tag{15}$$

It can be noted that the $H'(i, j)$ expressed by (13) is the element of matrix $\mathbf{H}$ with the row i and column j. It is the channel response from the beginning state q_j to the ending state q_i. So the matrix $\mathbf{H}$ includes the all frequency responses between any two points, together with the reflectometry information at any port.

$\mathbf{H}(i, j)$ is the frequency domain characteristic of channel from transmitter T_t to receiver T_r. In order to get the time-domain information, an inverse Fourier transform can be done to the sequence as follows

$$\rho_{ij}(t) = \text{IFT}(\mathbf{H}(i, j)) \tag{16}$$

It is worth to be noted that, if the transmitter T_t and receiver T_r are the same point, it is the situation of reflectometry, and $\rho_{ij}(t)$ is the reflectometry response, i.e., TDR. If they are different points, $\rho_{ij}(t)$ is the impulse response of the channel, i.e., $h_{ij}(t)$. Although there are some differences between two cases, $\rho_{ij}(t)$ contains the structural information of network topology.

2.4. Reconstruction of Network Topology from Reflectometry Responses by SVM

In the TDR curve, horizontal axis is the time $T = [t_1, t_2, \ldots, t_n]$, vertical axis is the corresponding reflection coefficient $\Gamma = [\Gamma_1, \Gamma_2, \ldots, \Gamma_n]$, both of them are discrete sequences. It is worth to note that the time sequence has the same interval Δt (i.e., $t_i = t_1 + \Delta t \times (i - 1)$). Generally, there are several pulses caused by the impedance discontinuities in the curve, each point (i.e., device, terminal, and fault) in the cable network is related to a pulse Γ_z in the TDR trace. For example, there is a sample TDR curve shown in Fig. 2. The first pulse at t_{51} is generated by point C_1, the second pulse at t_{136} is generated by T_2, and the third pulse at t_{177} is generated by the fault F_1. So, if the right corresponding pulse for each point can be distinguished, the fault network can be reconstructed by the estimated lengths of cables. For each point in the network, how to find the corresponding pulse's index z in $[\Gamma_1, \Gamma_2, \ldots, \Gamma_n]$ is an inversion problem, and can be solved by the SVM. Take the location of fault

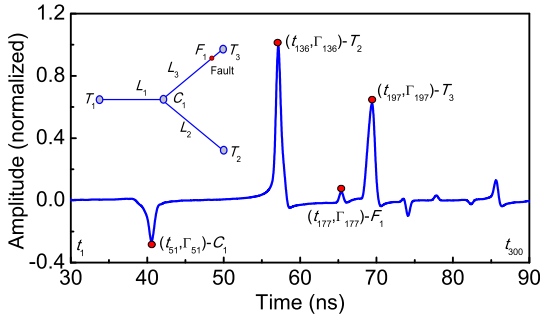

Figure 2. A sampled TDR curve at the port T_1.

F_1, for example, if the reflection coefficient $\Gamma = [\Gamma_1, \Gamma_2, \ldots, \Gamma_{300}]$ are obtained by measurement on actual network, the location of fault can be acquired by calculating the indexes related to C_1 and F_1 in Γ, i.e., 51 and 177, then the location of fault F_1 can be calculated as $(t_{177} - t_{51}) \times v = 126\Delta tv$, and the length of cable L_1 is $51\Delta tv$, where v is the velocity of propagation.

SVM maps the input data into a high dimensional feature space and constructs a linear regression function therein. The SVM model is given N training data $\mathbf{T} = \{(\mathbf{X}_i, y_i)\}_{i=1}^{N} \in \mathbf{R}_m \times \mathbf{R}$ where $\mathbf{X}_i$ is the input vector to the SVM model and y_i is the actual output value, from which it learns an input-output relationship. In our model, the input $\mathbf{X}_i$ is the reflection coefficient $\Gamma = [\Gamma_1, \Gamma_2, \ldots, \Gamma_m]$, the output is the corresponding index z (with Γ_z is the pulse caused by the estimated point). The SVM model can be expressed as follows:

$$y = f(\mathbf{X}) = \sum_{i=1}^{N} \omega_i \phi_i(\mathbf{X}) + b = \mathbf{W}^T \varphi(\mathbf{X}) + b \tag{17}$$

Equation (17) is a nonlinear regression model because the resulting hyper-surface is a nonlinear surface hanging over m-dimensional input space. The nonlinear function is learned using a linear learning machine of which the learning algorithm minimizes a convex functional. The coefficients $\mathbf{W}$ and b are the support vector weight and bias that estimated by minimizing the following regularized risk function:

$$R(\mathbf{W}) = 1/2\mathbf{W}^T\mathbf{W} + C\sum_{i=1}^{N} |y_i - f(\mathbf{X})|_\varepsilon \tag{18}$$

where

$$|y_i - f(\mathbf{X})|_\varepsilon = \max\{0, |y_i - f(\mathbf{X})| - \varepsilon\} \tag{19}$$

Constant C is known as a regularization parameter. The regularization parameter determines the trade-off between the approximation error and the weight vector norm. Insensitive loss parameter ε identifies allowed minimum fitting error of the learning machine, which affects the number of support vector (SV). The constants C and ε are user-specified parameters.

The foregoing regularized risk function is converted into the following constrained risk function:

$$R\left(\mathbf{W},\xi,\xi^*\right) = 1/2\mathbf{W}^T\mathbf{W} + C\sum_{i=1}^{N}\left(\xi+\xi^*\right) \tag{20}$$

Subject to the constraints

$$\begin{cases} y_i - \mathbf{W}^T\varphi(\mathbf{X}) - b \le \varepsilon + \xi_i \\ \mathbf{W}^T\varphi(\mathbf{X}) + b - y_i \le \varepsilon + \xi_i^*, \quad i = 1,2,\ldots,N \\ \xi_i, \xi_i^* \le 0 \end{cases} \tag{21}$$

The parameters ξ_i and ξ_i^* are the slack variables that represent the upper and lower constraints on the outputs of the system, respectively, and are positive values. The constrained optimization problem of (20) can be solved by its Lagrange dual:

$$\max_{\alpha_i,\alpha_i^*}\left\{ -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\left(\varphi(\mathbf{X}_i)\cdot\varphi(\mathbf{X}_j)\right) + \sum_{i=1}^{N}\alpha_i(y_i-\varepsilon-\xi_i) - \alpha_i^*(y_i-\varepsilon+\xi_i)\right\} \tag{22}$$

with constraints

$$\begin{cases} \sum_{i=1}^{N}(\alpha_i-\alpha_i^*) = 0 \\ 0 \le \alpha_i, \alpha_i^* \le C, \quad i = 1,2,\ldots,N \end{cases} \tag{23}$$

The kernel function $K(\mathbf{X}_i,\mathbf{X}) = \varphi^T(\mathbf{X}_i)\varphi(\mathbf{X})$ based on the Mercer condition is applied within the formulation of this quadratic programming problem. Solving (22) with constrains Equation (23) determines the Lagrange multipliers α, α^*, and the regression function of (17) becomes

$$y = f(\mathbf{X}) = \sum_{i=1}^{N}(\alpha_i-\alpha_i^*)K(\mathbf{X}_i,\mathbf{X}) + b \tag{24}$$

The corresponding Karush-Kuhn-Tucker complementarity conditions are

$$\begin{cases} \alpha_i \left(\mathbf{W}^T \varphi(\mathbf{X}) + b - y_i - \varepsilon - \xi_i\right) = 0 \\ \alpha_i^* \left(y_i - \mathbf{W}^T \varphi(\mathbf{X}) - b - \varepsilon - \xi_i^*\right) = 0 \\ (C - \alpha_i)\xi_i = 0, \qquad\qquad i = 1, 2, \ldots, N \\ (C - \alpha_i^*)\xi_i^* = 0 \end{cases} \tag{25}$$

Therefore, the support vectors are points where exactly one of the Lagrange multipliers is greater than zero. In the current research, the radial basis function (RBF) is used, i.e.,

$$K(\mathbf{X}_i, \mathbf{X}) = \exp(-\gamma \|\mathbf{X}_i - \mathbf{X}\|^2) \tag{26}$$

where γ is the kernel parameter, and decides the distribution complexity of sample data in high-dimensional feature space, $\mathbf{X}_i$ is the input vector to the SVM model.

The SVM optimization problem can be solved by dual formulation using many special-purpose solvers [17, 18]. One of the most commonly used solvers is LIBSVM [19]. The computational complexity of training nonlinear SVMs with LIBSVM has been reported to be quadratic in a number of training samples [20]. Before running LIBSVM code, three important hyperparameters should be identified: error penalty C, insensitive loss parameter ξ, and kernel parameter γ in SVM. In this paper, the leave-one-out cross-validation approach is adopted to infer values of the hyperparameters [21].

Using the generated datasets to train SVM, a certain interesting range which contains the important reflected pulses is chosen, such as time from t_m to t_n. So the inputs of the SVM is the corresponding reflection coefficient sequence $\Gamma = [\Gamma_m, \Gamma_{m+1}, \ldots, \Gamma_n]$, the output is the index z (where the corresponding Γ_z is related to a special point, such as C_1 in Fig. 2, and the number of inputs is $n - m + 1$. Generally, there is only one output in a SVM model, so that the number of SVM models for a cable network depends on the number of estimated points, such as C_1, T_1, T_2 and F_1 in Fig. 2.

The input/output database are randomly divided into two different sets: training set (80% of all samples) and testing set (20% of all samples). The training set is used to optimize the values of the hyperparameters and establish the best-performance SVM model; the testing set is employed to evaluate the generalization capability of the model. Based on the displacement error $e = y_{\text{model}} - y$, the mean absolute error (MAE) and root-mean-square error (rmse) of this model

are defined as follows:

$$\text{MAE} = \frac{1}{N_t}\sum_{i=1}^{N_t}|e_i| \tag{27}$$

$$\text{rmse} = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}e_i^2} \tag{28}$$

where N_t is the number of testing data set. Therefore, these two quantitative metrics are used to indicate the performance of the SVM model.

After being trained correctly, the SVM can effectively distinguish the pulses in the TDR curve and calculate the length of each cable from the measured reflectometry trace. Meanwhile, multiple faults on the network can be located accurately.

To better understand the method, its flow diagram is shown in Fig. 3. There are about five steps to implement the method: Step 1: The topology of analyzed network need to ascertain, including the type of each component. However, it is not necessary to know the accurate length of each cable, as they can be calculated by the SVM from the measured TDR. Step 2: The model is built for each component in the network, including the coaxial cable, splitter, tap and so on. Step 3: According to the topology of network, different lengths of branches and different locations of multi-faults are set to the distribution network, the range must be limited. Then, thousands of network samples, with the same topology but different configuration, can be generated. Therefore, thousands of TDR curves can be calculated by the State-transition matrix. Step 4: The TDR curves are used to train the SVMs;

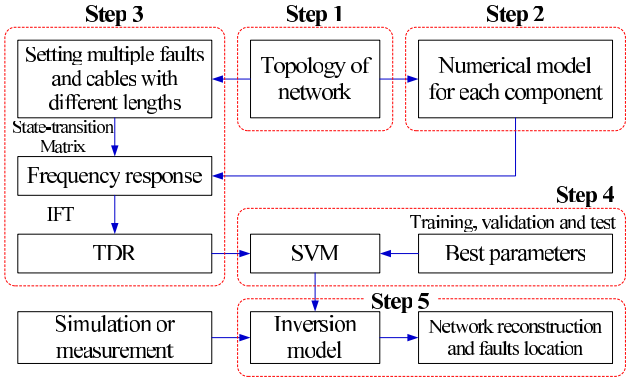

Figure 3. The flow diagram of this proposed method.

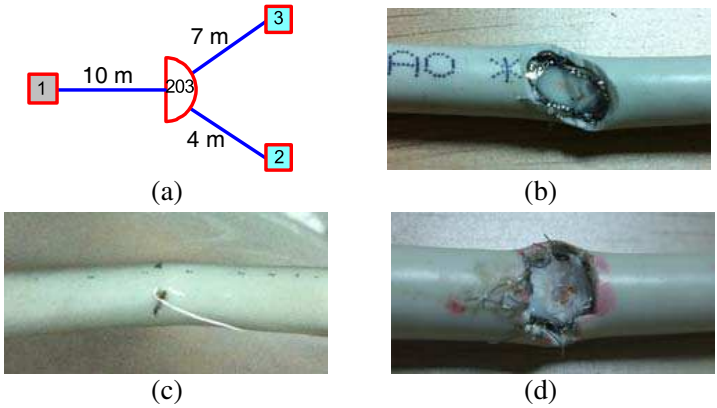

Figure 4. A network sample in our laboratory. (a) Healthy network. (b) Seriously damaged insulation. (c) Water infiltration (with salt). (d) Hard fault.

meanwhile the leave-one-out cross-validation approach is adopted to infer values of the hyperparameters. Step 5: After being trained correctly, the SVM can effectively calculate the length of each cable and extract the network topology from the measured reflectometry trace, locate faults on the network which belong to the topology as well.

3. NUMERICAL AND EXPERIMENTAL RESULTS

To verify the performance of proposed SVM model, three coaxial cable networks have been considered. As mentioned above, there are two types of faults: soft faults and hard faults. Both of them will be discussed as follows.

3.1. Location of Soft Faults

The soft faults are caused by discontinuities of impedance, such as the damaged insulation and water infiltration. The first simple analyzed topology is shown in Fig. 1. First, a network sample is set up in our laboratory, and different types of fault are introduced on cable L_3, such as damaged insulation, water infiltration and hard fault. Their photos are shown in Fig. 4. Then, the scattering parameters of each fault network in the frequency domain ranging from 1 MHz to 1.2 GHz are measured. Finally, the TDR can be obtained by IFT, and will be compared with the simulated results.

When a segment of length l_F and width w_F is chafed on a coaxial cable, the chafed segment can be regarded as having a constant (l_F

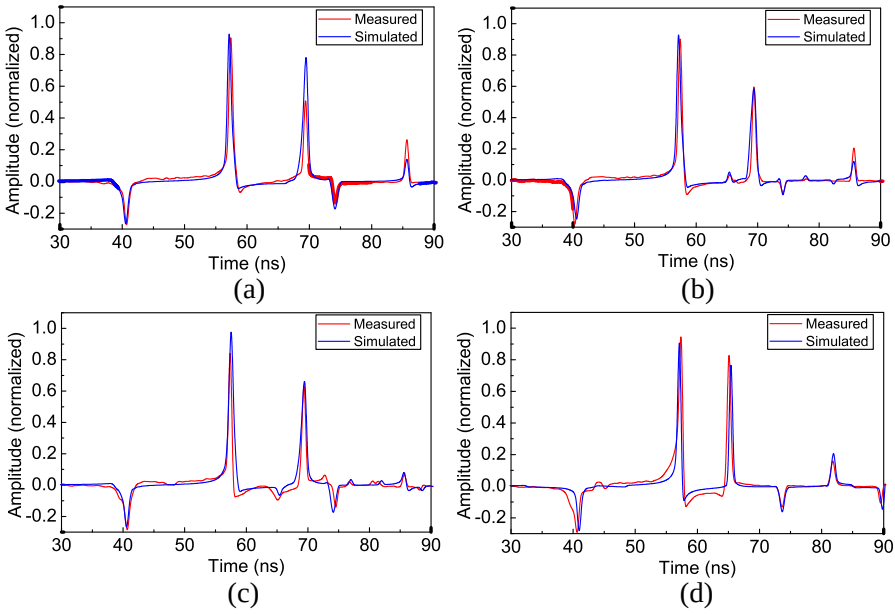

Figure 5. The simulated and measured TDR at port 1. (a) Healthy network. (b) Damaged insulation. (c) Water infiltration. (d) Hard fault.

and w_F-independent) impedance Z_F [6], and $Z_F > Z_0$ (with Z_0 is the characteristic impedance of coaxial cable). The theoretical underpinnings and the numerical implementation of this approach are presented in [22]. For the water infiltration (with salt), Z_F is always less than Z_0; meanwhile, in terms of the hard faults (break point), Z_F is treated as infinite.

Figure 5 presents the simulated and measured results of TDR response at the port 1. It is very clear that all the results are very similar in terms of location and magnitude of the main peaks. The existing error is mainly caused by the connector between VNA and network.

It is worth to point out that, compared to the healthy network, the soft faults only cause only small changes to the TDR curve (In fact, the faults introduced in Fig. 4 are very grievous. For instance, plenty of salt is added into the water for the infiltration). These changes are always below the level of error shown in the plot, especially for the complex network. Hence, it is difficult to detect the soft faults at the port which is far away from the faulty point, such as port 1 in Fig. 4. In this case, soft faults can be detected at the nearest port (in fact,

it is the terminal where the fault happened), such as the port 3. The fault point and the tested port are in the same cable, there is no device between them. So the power of the reflected signal will be relatively larger and can be detected by the instrument. The simulated and measured TDR curves at port 3 are shown in Fig. 6.

It can be seen from the Fig. 6, changes caused by the fault can be displayed in the TDR curves. In fact, this case is considered as the fault location on single cable. Now, the SVM model is used to locate soft faults at the nearest port. Soft faults with different impedance and different location are introduced on the cable L_3. Then thousands of TDR curves at port 3 can be calculated by the State-transition matrix. Some are shown in Fig. 7. Just like the "TDR1", the soft fault is 5 m away from port 3, and its impedance is 30 Ω.

Two SVMs can be designed to diagnose the soft faults: SVM_1 is to locate the fault; SVM_2 is to estimate the impedance of fault. These thousands of TDR curves are used to train the SVMs. In the plot, the time interval Δt chosen in the study is 0.2 ns. Our interesting time range is from 0 to 32 ns, there are 160 points in the TDR curve (i.e., (t_i, Γ_i), $i \in [1, 2, \ldots, 160]$). The input vector of the two SVMs is

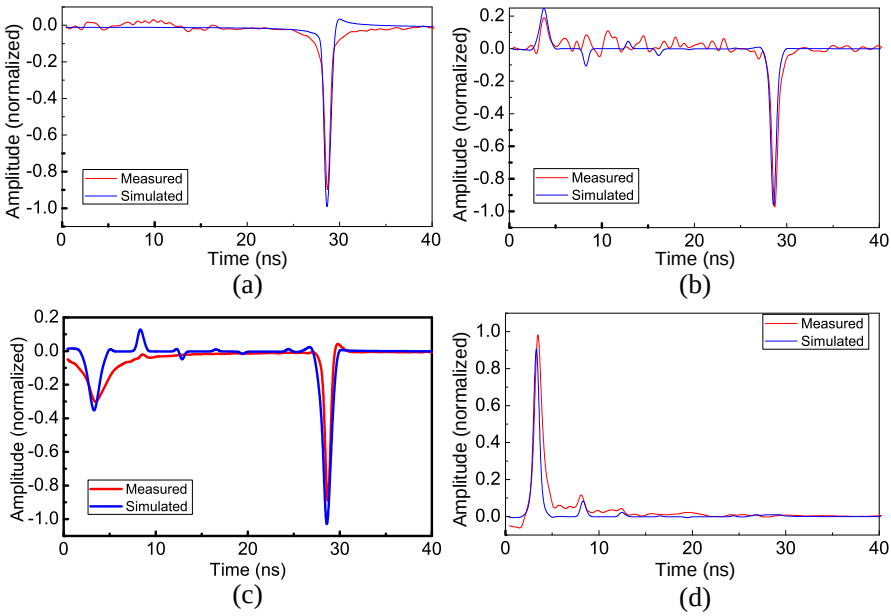

Figure 6. The simulated and measured TDR at port 3. (a) Healthy network. (b) Damaged insulation. (c) Water infiltration. (d) Hard fault.

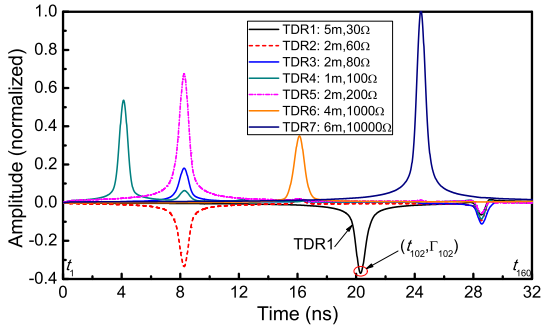

Figure 7. Simulated TDR curves with different fault locations and impedances.

$[\Gamma_1, \Gamma_2, \ldots, \Gamma_{160}]$, the output of SVM_1 is the index z (i.e., $z = 102$ for "TDR1"), the output of SVM_2 is the impedance R (i.e., $R = 30\,\Omega$ for "TDR1").

1000 samples included input datasets (TDR response) and output datasets (fault location and impedance) have been created by the aforementioned model, 800 random samples are used for training, 200 for the testing. Then hyperparameters of the two SVMs are obtained by the leave-one-out cross-validation method ($C = 1.3192$, $\varepsilon = 0.5$, and $\gamma = 0.152$ for SVM_1; $C = 6.964$, $\varepsilon = 0.75$, and $\gamma = 0.658$ for SVM_2).

Once the training process is completed, 10 random samples from the testing datasets are used to evaluate the capability of these generated SVMs. It is worth to note that the estimated z should be changed to the fault location as $t_z \times v$, where v is the velocity of propagation. Fig. 8 shows the comparison between the actual and simulated location and impedance of fault on the branch. Concerning the predicting errors, the calculated MAE and rmse of fault locations are 0.079 m and 0.088 m respectively. Because the range of fault impedance is too wide, the mean relative error is chosen as metric, and the calculated value is 11.5%. These results illustrate clearly the good performance of SVM models to diagnose soft faults.

Although the generated SVMs exhibit a good performance in soft fault diagnosis, they only detect faults at the nearest port, but not the headmost port. How to diagnose and locate soft faults on complex network at the headmost port, is a difficult but important issue, and will be addressed in the future research. In this paper, we primarily focus on the location of hard faults on complex network.

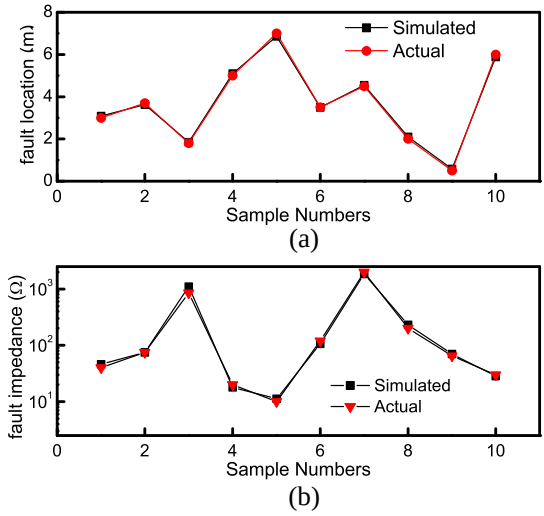

Figure 8. (a) Actual and simulated fault locations. (b) Actual and simulated fault impedance.

3.2. Location of Hard Faults

When a hard fault locates on a branch, comparing to the healthy network, the topology of the faulty network does not change. The difference is only the length of branch. In other term, for the location of hard faults, It is only needed to estimate the later length of each branch from the measured TDR trace, and reconstruct a network with new configuration as the faulty network. Meanwhile, in most residential areas, the topology of the cable distribution network is the same, the difference is only the length of cable. In practice thousands of network samples, which belong to the same topology but with different lengths of branches, are used to train the SVM. After being trained correctly, the SVM can estimate length of each cable and locate faults for the network which belongs to the analyzed topology.

In order to get a clearer understanding of the SVM model, a more complex topology of network shown in Fig. 9 is studied. The type of the 2-way tap is "208", the 2-way splitter is "203".

To simplify, it is supposed that the hard faults only locate on branches (L_2, L_3, L_4 and L_5), and the main cable are $L_1 = 20\,\text{m}$, $L_6 = 10\,\text{m}$. Different lengths are set to branches (the range is from 0 to 6 m), and then thousands of TDR curves can be obtained. According to the number of varied branches, the number of SVMs is limited to four. It is worth to point out that this network is double-symmetrical,

i.e., L_2-L_3, and L_4-L_5. Hence, when the thousand TDR curves are used to train the SVMs, a flag parameter is added to input data: for SVMs of L_2 and L_3, if $L_2 \geq L_3$, then $\Gamma_{\text{flag}} = 1$, otherwise $\Gamma_{\text{flag}} = -1$; and, it is the same way for L_4 and L_5.

After training correctly, 5 samples are used to test the model, and the results are shown in Fig. 10. The MAEs are 0.03 m, 0.04 m, 0.03 m, 0.04 m, respectively. So it demonstrates that the SVM model also has a decent ability to locate multiple faults for the complex network.

The five evaluated samples are from the testing datasets. Like the

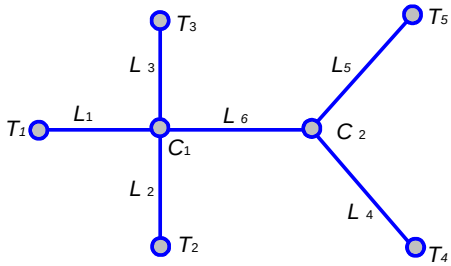

Figure 9. A topology of network with a tap and a splitter.

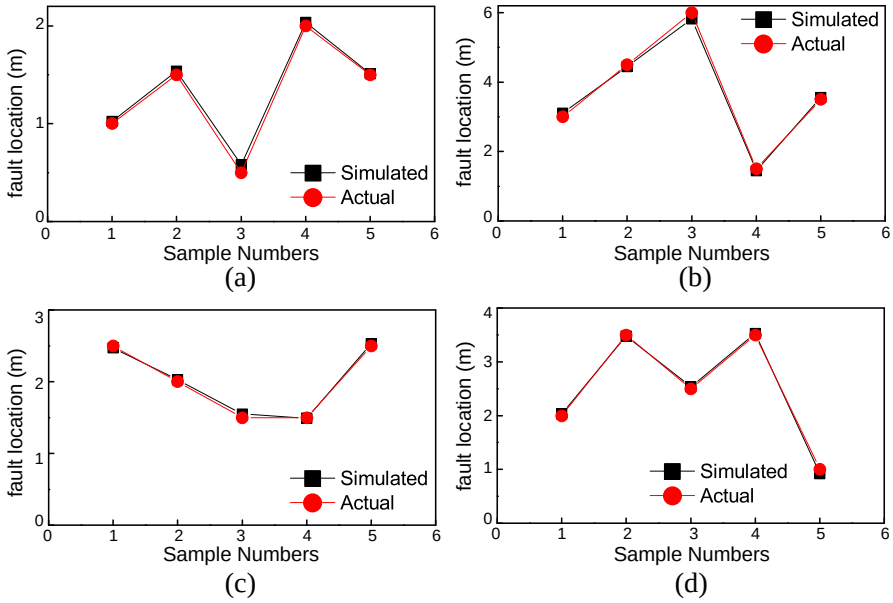

Figure 10. Actual and simulated faults location. (a) Branch L_2. (b) Branch L_3. (c) Branch L_4. (d) Branch L_5.

training datasets, all of them are generated by the State-transition matrix. Therefore, it is necessary for this SVM model to locate faults from the measured TDR trace. A network which belongs to the topology of network in Fig. 9 is set up in our laboratory, and two faults are introduced on the branches L_2 and L_5. It is depicted in Fig. 11(a). The measured TDR trace is shown in Fig. 11(b), and will be used to validate the generated SVM.

As mentioned about, this is a double-symmetrical network, when using the trained SVMs to reconstruct fault network, the added flag parameter Γ_{flag} should be set. At first, supposing that $L_2 > L_3$ and $L_4 > L_5$ in the faulty network, so that, the $\Gamma_{\text{flag}} = 1$ for SVMs of L_2 and L_3, and it is the same way for SVMs of L_4 and L_5. Then, the regressive results are obtained by the SVM: $L_2 = 4.58\,\text{m}$, $L_3 = 1.96\,\text{m}$, $L_4 = 3.07\,\text{m}$ and $L_5 = 1.95\,\text{m}$. After that, the results are combined with the original network to analyze our supposal. For example, the original length of L_2 is 2 m in the healthy network, L_2 must be less than 2 m in faulty network, hence it can be inferred that: $L_2 = 1.96\,\text{m}$, $L_3 = 4.58\,\text{m}$. Finally, the simulated lengths of branches are $L_2 = 1.96\,\text{m}$, $L_3 = 4.58\,\text{m}$, $L_4 = 3.07\,\text{m}$ and $L_5 = 1.95\,\text{m}$. The predicting errors can be calculated that the MAE $= 0.06\,\text{m}$ and rmse $= 0.06\,\text{m}$. The existing error in this model is mainly caused by three reasons. First, ambiguity in the velocity of propagation is proportional to ambiguity in the location of the fault, and v may vary with the ambiance, especially the temperature. Second, there is some error in measuring the true lengths of cables. Third, the splitter and tap will led to time delay, which is not considered in this model.

Figure 12 shows an actual distribution network for two buildings. Each building has six floors and each floor has two users. This network

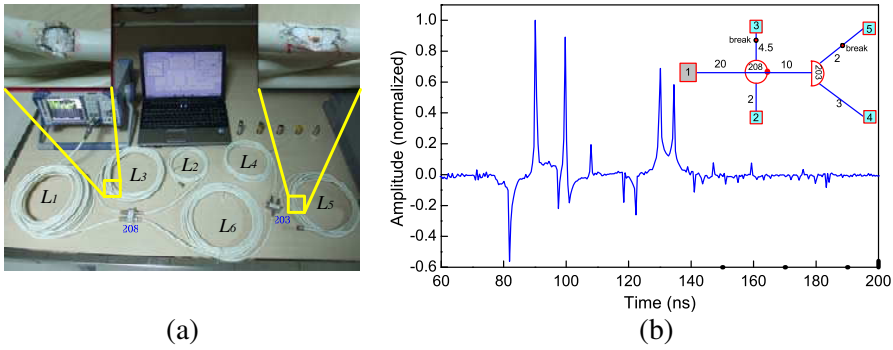

Figure 11. (a) Cable network with two faults. (b) Measured TDR response.

is much more symmetry, to simplify, faults are only introduced on cable L_1, L_3 and L_5. The measured TDR response at port 0 is shown in Fig. 13. After SVM is trained correctly, this TDR trace is used to estimate the fault locations, and the results are shown in Fig. 14. The predicting errors are MAE = 0.13 m and rmse = 0.14 m. The results demonstrate that the proposed SVM model is also effective to locate faults on actual complex network.

The SVM model can locate faults from the measured TDR trace. But for the symmetrical network, especially the complex network, it is still not clear which branch the fault lies on. In this case, more information is required to determine the real configuration of the faulty network, such as the original lengths of cables in the healthy network,

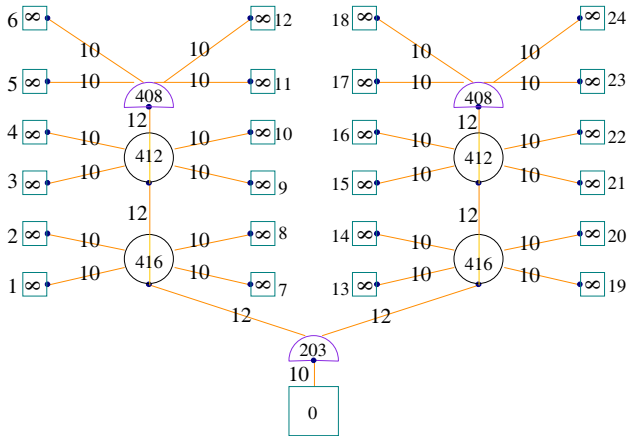

Figure 12. A network sample in our laboratory.

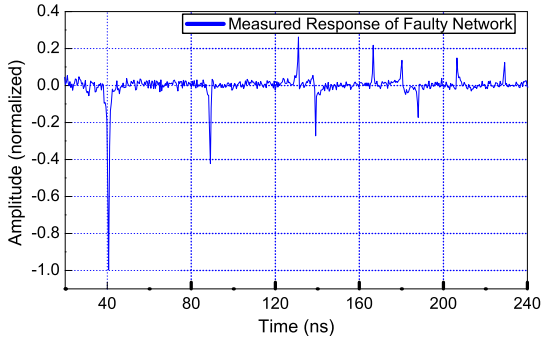

Figure 13. The measured TDR of the network shown in Fig. 12.

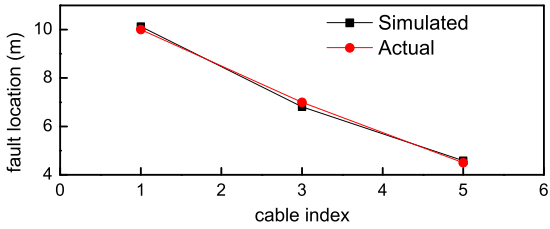

Figure 14. Actual and simulated faults location.

the state of all terminals or more TDR responses at other terminals. Meanwhile, it should be pointed out that, the inputs of SVM are the all points in TDR curve, for complex network, the number of inputs may be very large. Although the SVM exhibits the major advantage of input-dimension independence, it needs a large number of training datasets. So in the future work, before using the SVM to solve the inversion problem, the TDR trace should be analyzed at first, such as feature extraction and principal component analysis, to reduce the number of inputs.

4. CONCLUSION

Only need to know the topology of distribution network, a method, allowing the faulty network to be reconstructed by estimating the lengths of branches, is proposed in the paper. This method is based on the State-transition matrix model and SVM. The state-transition matrix can generate the TDR response at any terminals, and simulate the frequency and impulse response between any two points. This is helpful to better understand the channel behavior of complex cable network. How to locate faults from the measured TDR trace is the inversion problem and can be solved by the SVM model. The training datasets required by SVM are created by the State-transition matrix. Once the training process is completed, the formed SVM can reconstruct the faulty network by finding the length of each branch. Experimental results demonstrate that the SVM model has a good capability to locate faults on complex network, especially for multiple-fault network. When the TDR curve is measured at the terminal port, this approach can diagnose the soft faults. Furthermore, this approach can be exploited in embedded system. However, it should be pointed out that, for the symmetrical network, it is still hard for this model to determine which branch the fault lies on. Only knowing the TDR curve at one port is not enough to reconstruct the faulty network. This puzzle problem will be addressed in the future study.

ACKNOWLEDGMENT

This work was fully supported by the National Natural Science Foundation of China under Grant 61107051, the National High Technology Research and Development Program of China under Grant SS2012AA010407, and the Major Program of Natural Science Foundation of Hubei Province under Grant 2011CDA114.

REFERENCES

1. Boyd, E., H. Elbakoury, M. Hajduczenia, and A. Liu, "EPON over Coax (EPoC)," *IEEE Commun. Mag.*, Vol. 50, No. 9, 88–95, 2012.
2. IEEE 802.3 Working Group, 2012, http://www.ieee802.org/3/ep-oc.
3. Lelong, A., L. Sommervogel, N. Ravot, and M. Carrion, "Distributed reflectometry method for wire fault location using selective average," *IEEE Sens. J.*, Vol. 2, 300–310, 2010.
4. Cataldo, A., G. Cannazza, E. De Benedetto, and N. Giaquinto, "Experimental validation of a TDR-based system for measuring leak distances in buried metal pipes," *Progress In Electromagnetics Research*, Vol. 132, 71–90, 2012.
5. Kwak, K. S., T. Choe, J. Park, and T. Yoon, "Application of time-frequency domain reflectometry for measuring load impedance," *IEICE Electronics Express*, Vol. 5, 107–113, 2008.
6. Schuet, S., D. Timucin, and K. Wheeler, "A model-based probabilistic inversion framework for characterizing wire fault detection using TDR," *IEEE Trans. Instrum. Meas.*, Vol. 60, 1654–1663, 2011.
7. Pourahmadi-Nakhli, M. and A. A. Safavi, "Path characteristic frequency-based fault locating in radial distribution systems using wavelets and neural networks," *IEEE Trans. Power Del.*, Vol. 26, 772–781, 2011.
8. Vakula, D. and N. V. S. N. Sarma, "Using neural networks for fault detection in planar antenna arrays," *Progress In Electromagnetics Research Letters*, Vol. 14, 21–30, 2010.
9. Meng, J., Y. Gao, and Y. Shi, "Support vector regression model for measuring the permittivity of asphalt concrete," *IEEE Microw. Wirel. Co.*, Vol. 17, No. 12, 2007.
10. Zhang, Y. and L. Wu, "An Mr brain images classifier via principal component analysis and kernel support vector machine," *Progress In Electromagnetics Research*, Vol. 130, 369–388, 2012.
11. Thukaram, D., H. P. Khincha, and H. P. Vijaynarasimha,

"Artificial neural network and support vector machine approach for locating faults in radial distribution systems," *IEEE Trans. Power Del.*, Vol. 20, No. 2, 710–721, 2005.

12. Angiulli, G., D. De Carlo, G. Amendola, E. Arnieri, and S. Costanzo, "Support vector regression machines to evaluate resonant frequency of elliptic substrate integrate waveguide resonators," *Progress In Electromagnetics Research*, Vol. 83, 107–118, 2008.
13. Ni, J., L. Ren, C. Zhang, and S. Yang, "Abrupt event monitoring for water environment system based on KPCA and SVM," *IEEE Trans. Instrum. Meas.*, Vol. 61, 980–989, 2012.
14. Wu, Y., Z. X. Tang, B. Zhang, and Y. Xu, "Permeability measurement of ferromagnetic materials in microwave frequency range using support vector machine regression," *Progress In Electromagnetics Research*, Vol. 70, 247–256, 2007.
15. Chen, W. Y. and K. Kerpez, "Coaxial cable distribution plant performance simulation for interactive multimedia TV," *Global Telecommunications Conference*, 173–177, 1995.
16. Zimmermann, M. and K. Dostert, "A multipath model for the powerline channel," *IEEE Trans. Commun.*, Vol. 50, 553–559, 2002.
17. Cristianini, N. and J. S. Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, London, 2000.
18. Tan, C. P., J. Y. Koay, K. S. Lim, H. T. Ewe, and H.-T. Chuah, "Classification of multi-temporal SAR images for rice crops using combined entropy decomposition and support vector machine technique," *Progress In Electromagnetics Research*, Vol. 71, 19–39, 2007.
19. Bermani, E., A. Boni, A. Kerhet, and A. Massa, "Kernels evaluation of SVM based-estimators for inverse scattering problems," *Progress In Electromagnetics Research*, Vol. 53, 167–188, 2005.
20. Bottou, L., O. Chapelle, D. DeCoste, and J. Weston, Editors, *Large Scale Kernel Machines*, MIT Press, Cambridge, MA, 2007.
21. Cawley, G. C., "Leave-one-out cross-validation based model selection criteria for weighted LS-SVMs," *International Joint Conference on Neural Networks, IJCNN*, 1661–1668, 2006.
22. Kowalski, M., "A simple and efficient computational approach to chafed cable time-domain reflectometry signature prediction," *Proc. Annu. Rev. Progress ACES Conf.*, 2009.