

REVERSE ENGINEERING OF SIGNAL ACQUISITION CHAINS USING THE THEORY OF SAMPLING SIGNALS WITH FINITE RATE OF INNOVATION

Thirapiroon Thongkamwitoon, Hani Muammar, and Pier Luigi Dragotti

Electrical and Electronic Engineering Department, Imperial College London, United Kingdom

ABSTRACT

This paper presents a novel theoretical framework for the reverse engineering of signal acquisition chains. We investigate how signals are transformed through a chain of signal acquisition and reconstruction stages. The signals, at different stages in the chain, are modelled using the theory of sampling signals with Finite Rate of Innovation (FRI). The model allows us to determine the chain structure and corresponding acquisition history from unknown query signals and to retrieve important parameters relating to the acquisition chain.

Index Terms— Digital Forensics, Reverse Engineering, Signal Acquisition, Acquisition Chain, Signals with FRI

1. INTRODUCTION

Nowadays digital signals can be captured and manipulated many times during their lifetime. For example, a signal may go through multiple A/D and D/A conversions, it might be encoded and decoded several times or it may be edited. Consequently, it becomes important to be able to recover the processing history associated with a digital signal including the properties of the device used to acquire the signal, the compression algorithms involved and the post-processing techniques used. Reverse-engineering a complex chain of operators is relevant not only in forensics but in areas such as quality assessment and signal enhancement.

Over the last few years, the digital forensic community has been highly active in developing novel approaches to recovering the processing history of a digital signal. In one example, Swaminathan et al. [1] introduce the concept of component forensics based on footprints left from different components of acquisition devices. The authors propose a method to predict the components in an information processing chain from query data using pattern classification theory. Forensics, for a cascade of digital sampling operations, was first investigated in [2]. The paper develops a resampling detection algorithm based on the correlation of periodic artefacts

from resampling with those of neighboring pixels. It was further studied in the spectral domain using artifacts from the residue of a local linear predictor [3]. In [4], the roles of pre-filtering and cyclostationarity of signals are studied in order to improve the performance of the resampling detection algorithm. The statistical properties of periodic artefacts left by interpolation are examined in [5]. In previous work [6] the authors investigate how edge profiles in images are transformed by the image acquisition chain.

An understanding of when the processing footprint left by an operator is completely removed by other operators or when a processing chain becomes too complex to be completely retrieved is, generally, of interest. It is anticipated that some operators may accidentally completely remove some footprints related to previous processing. For example, low rates of compression can remove many fine features from a signal. It would, therefore, be nice to obtain an understanding of the circumstances in which a chain can always be precisely recovered and when two processing chains, overall, produce the same artefacts and, therefore, become indistinguishable.

An attempt at addressing this question is presented in this paper. We focus only on multiple A/D and D/A conversions and, specifically, consider the case where the digital signal is recaptured at most once. The A/D and D/A operators are modelled using generalized sampling theory [7], [8] and the theory of sampling FRI signals [9], [10] is used to provide some precise answers. We model the original signal as a step function because in 2-D this feature would correspond to a straight edge (a feature that is abundant in natural images) and we study the conditions under which we can detect recapture and identify the parameters of the original A/D and D/A operators. The major contribution of our framework is a complete end-to-end analytical method for reverse engineering of acquisition chains.

The paper is organized as follows: Section 2 presents the problem model, priori conditions, and associated problem statements. The background of the theory of sampling signals with FRI is provided in Section 3. In Section 4 we show how to employ FRI sampling theory to provide sufficient conditions for reacquisition detection and parameter retrieval of acquisition chains. We also show with a counter example that when such conditions are not satisfied two acquisition chains produce the same output. Finally, we conclude in Section 6.

This work is supported by the REWIND Project, funded by the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, FET Open grant number:268478. Thirapiroon Thongkamwitoon is supported by The Office of NBTC and the Royal Thai Scholarship, THAILAND.

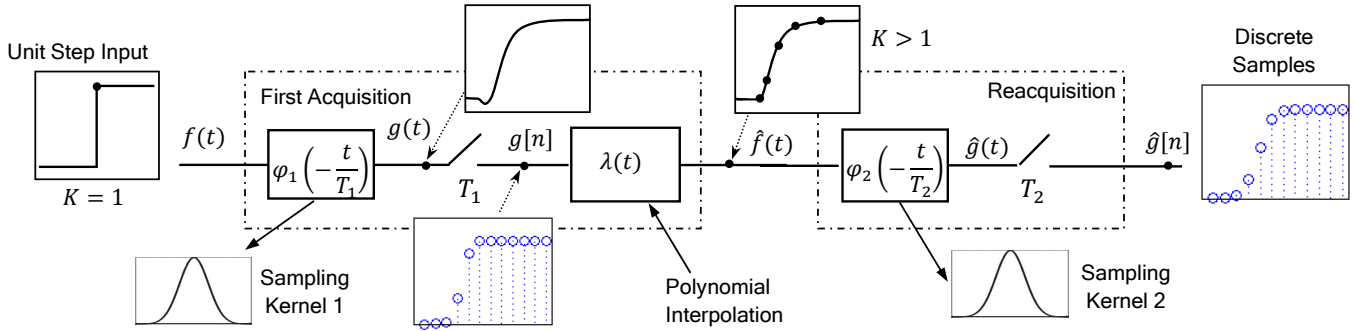


Fig. 2. Problem statement diagram for signals with FRI in the chain of signal acquisition.

2. PRELIMINARIES AND PROBLEM SETUP

We use the generalized sampling theory to model the sampling and reconstruction of a signal $x(t)$. This process is depicted in Figure 1 where the signal $x(t)$ is filtered before being uniformly sampled. This leads to the measurements $y[n] = \langle x(t), \varphi(t/T - n) \rangle$, where the sampling kernel is the time reversed and scaled version of the filter's impulse response $h(t)$. Reconstruction is achieved using the linear filter $\lambda(t)$ which yields $\hat{x}(t) = \sum_{n \in \mathbb{Z}} y[n] \lambda(t/T - n)$. We assume that $\lambda(t)$ is a polynomial spline or a MOMS function [11] of order R , and therefore $\hat{x}(t)$ is a piecewise polynomial function of maximum order R . We also assume $\hat{x}(t) \neq x(t)$.

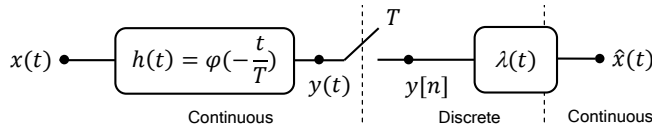


Fig. 1. Classical signal acquisition and reconstruction model.

Next, we construct a study problem for the chain of acquisition and reconstruction as shown in the diagram in Figure 2. The chain contains sampling, reconstruction, and sampling modules aligned serially as shown in the diagram. The parameters $\varphi_1(t)$, T_1 , $\varphi_2(t)$, and T_2 , represent the sampling kernels and sampling periods of the first and second acquisition devices, respectively. With this chain structure, the problem conditions are set as follows:

- 1) The input signal is fixed as a box function $f(t) = u(t - t_1) - u(t - t_2)$, where t_1 and t_2 are the unknown locations of the unit step functions and $t_1 < t_2$. Initially, we consider $t_2 \rightarrow \infty$. Thus, the input $f(t)$ can be approximated by a step $u(t - t_s)$. The role of t_2 will be discussed in Section 4.
- 2) The type of interpolation is polynomial, with maximum degree R .
- 3) The second sampling kernel $\varphi_2(t)$ was introduced in [10] and has the special properties that it can reproduce polynomials or exponentials.

Given access to only a query digital signal $q[n]$, the key

questions are:

- a) What stages in the chain are the samples $q[n]$ from? That is, was $q[n]$ obtained by acquiring $f(t)$ directly with $\varphi_2(t)$ or was $q[n]$ the reacquired signal $\hat{g}[n]$ in Figure 2?
- b) In the case of reacquisition, how can we retrieve the following important parameters: i) the maximum order R of polynomial used for interpolation ii) the sampling period T_1 , and iii) the sampling kernel $\varphi_1(t)$?
- c) Under what conditions of, φ_2 , T_2 , can we solve (b)?

3. THE THEORY OF SAMPLING SIGNALS WITH FINITE RATE OF INNOVATION

A signal with FRI is defined as a signal which has a finite number of degrees of freedom per unit of time. Given a finite number of shifts t_k and amplitudes $\alpha_{k,r}$, a signal with FRI $x(t)$ can be described by known functions $\{f_r(t)\}_{r=0}^{R-1}$ as follows:

$$x(t) = \sum_{k \in \mathbb{Z}} \sum_{r=0}^{R-1} \alpha_{k,r} f_r(t - t_k). \quad (1)$$

Examples of signals with FRI include a stream of Diracs, a stream of differentiated Diracs and piecewise polynomial functions. With reference to the classical sampling diagram shown in Figure 1, we now consider a sampling scheme for signals with FRI. Let the input signal $x(t)$ be a train of K Diracs which is described by K pairs of free parameters: the locations t_k and the amplitudes a_k as follows:

$$x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k). \quad (2)$$

The input is then sampled with the sampling kernel $\varphi(t)$ with period T before discrete samples $y[n]$ are obtained. In this paper, we assume that the sampling kernel used is a function that can reproduce polynomials as described in [10]. For polynomial reproducing kernels there exists a set of coefficients $c_{n,p}$ such that:

$$\sum_{n \in \mathbb{Z}} c_{n,p} \varphi\left(\frac{t}{T} - n\right) = t^p; \quad p = 0, 1, 2, \dots, P. \quad (3)$$

Next, the moments τ_p of order p of the signal can be computed as follows:

$$\begin{aligned} \tau_p &= \sum_n c_{n,p} y[n] \stackrel{(a)}{=} \langle x(t), \sum_n c_{n,p} \varphi(t/T - n) \rangle \\ &\stackrel{(b)}{=} \langle \sum_{k=0}^{K-1} a_k \delta(t - t_k), \sum_n c_{n,p} \varphi(t/T - n) \rangle \\ &\stackrel{(c)}{=} \sum_{k=0}^{K-1} a_k t_k^p; \quad p = 0, 1, 2, \dots, P, \end{aligned} \quad (4)$$

where (a) follows from the linearity of the inner product while (b) and (c) are from equations (2) and (3) respectively. Once the moments τ_p ; $p = 0, 1, \dots, P$ and $P \geq 2K$ have been computed, the following Toeplitz matrix is constructed:

$$S = \begin{bmatrix} \tau_K & \tau_{K-1} & \cdots & \tau_0 \\ \tau_{K+1} & \tau_K & \cdots & \tau_1 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_P & \tau_{P-1} & \cdots & \tau_{P-K} \end{bmatrix}. \quad (5)$$

Note that, one can show [10] that S always has rank K (number of Diracs in $x(t)$) and that $x(t)$ is determined from the knowledge of the null space of S . Next, consider an input signal which is a piecewise polynomial signal with K pieces of maximum degree $R \geq 0$, that is:

$$x(t) = \sum_{k=1}^K \sum_{r=0}^R a_{k,r} (t - t_k)^r. \quad (6)$$

Clearly the $(R + 1)$ order derivative $x^{(R+1)}(t) = \frac{d^{(R+1)}x(t)}{dt^{(R+1)}}$ is given by a train of differentiated Diracs at the locations t_k as follows:

$$x^{(R+1)}(t) = \sum_{k=0}^{K-1} \sum_{r=0}^R r! a_{k,r} \delta^{(R-r)}(t - t_k). \quad (7)$$

We observe that $x^{(R+1)}(t)$ is an FRI signal. We also note that the finite difference $z^{(1)}[n]$ satisfies [10]:

$$\begin{aligned} z^{(1)}[n] &= y[n+1] - y[n] \\ &= \langle x(t), \varphi(t/T - n - 1) - \varphi(t/T - n) \rangle \\ &= \left\langle \frac{dx(t)}{dt}, \varphi(t/T - n) * \beta_0(t/T - n) \right\rangle, \end{aligned} \quad (8)$$

where β_0 is the box function. Therefore, the moments of the derivative of $x(t)$ are given by $\tau_p = \sum_n c_{n,p}^{(1)} z^{(1)}[n]$, where $c_{n,p}^{(1)}$ are the polynomial reproduction coefficients of (3) for the new kernel $\varphi(t) * \beta_0(t)$. The moments of the $R + 1$ derivative of $x(t)$ can be obtained similarly. Finally, it is again possible to show that the Toeplitz matrix S of the moments of $x^{(R+1)}(t)$ has rank proportional to the degrees of freedom of $x^{(R+1)}(t)$.

4. REACQUISITION DETECTION AND THE RETRIEVAL OF CHAIN PARAMETERS

We are given a query digital signal $q[n]$ and we want to determine whether this is the result of acquiring the unit step function, $f(t) = u(t - t_s)$, with $\varphi_2(t)$ or whether this is the result of reacquisition. An illustrative example of the two possible shapes of $q[n]$ is shown in Figure 3. In Figure 3(c) we show the case of a single acquisition of $f(t)$ shown in Figure 3(a), whereas Figure 3(d) shows a reacquired signal obtained after linear interpolation of (c) to yield 3(b) and sampling of 3(b) with $\varphi_2(t)$. We note that $g[n]$ and $\hat{g}[n]$ are hardly distinguishable yet they still contain all the information necessary to reverse engineer the acquisition chain as is shown next.

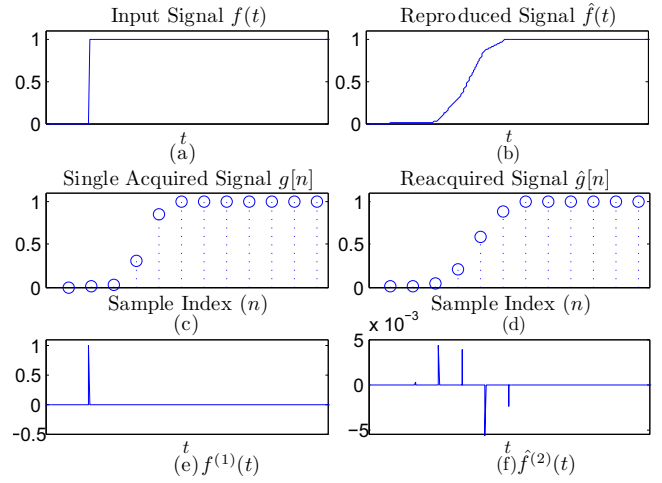


Fig. 3. Comparative plots of continuous input signals (a) step input $f(t)$ (b) reconstructed signal $\hat{f}(t)$ with corresponding discrete samples (c) $g[n]$ and (d) $\hat{g}[n]$ and reconstructed locations using FRI sampling theory (e) and (f) respectively.

The input signal $f(t)$ is a unit step function which is described by only one free parameter – the location of step t_s . When the input signal $f(t)$ is acquired, the observed samples $g[n]$ are distorted by the sampling kernel. All possible $g[n]$, however, are still determined by one free parameter. In contrast, $\hat{f}(t)$ is obtained by polynomial interpolation and is a polynomial function with discontinuities at multiple locations of period T_1 . The signal is a special case of FRI signals in (6).

We thereby use this principle to create an algorithm for reacquisition detection. We first aim to detect whether a query $q[n]$ was acquired singly or was reacquired. Since a step function is a piecewise polynomial of maximum degree $R = 0$, the moments are computed using a first order finite difference of the query as $\tau_p = \sum_n c_{n,p} q^{(1)}[n]$. The moments are then used to construct the Toeplitz matrix S . The matrix S of size 2×2 is sufficient for reacquisition detection since the matrix is always rank-deficient with rank = 1 if $q[n]$ was acquired from a step input. If, on the other hand, S is full rank, it means $q[n]$ was obtained by reacquisition.

When a query signal is detected as having been reacquired, the question arises of how important image chain parameters such as the sampling period T_1 , the interpolation function $\lambda(t)$, and the first sampling kernel φ_1 can be retrieved.

Firstly, the maximum order R of the polynomial interpolation function $\lambda(t)$ can be retrieved from the properties of FRI reconstructed signals. According to Section 3, piecewise polynomial functions of maximum order R are fully suppressed by differentiation by order $R + 1$. If we measure the number of degrees of freedom using Toeplitz matrix S , the matrix will be full rank until a finite difference of order $r \geq R + 1$ is applied to query samples $q[n]$. When $r = R + 1$, the matrix will be rank deficient with rank, K , equal to the number of K pieces of a piecewise polynomial function. Figure 4 summarizes the retrieval algorithm, using iterative finite difference and rank measurement, for order R until S is rank deficient.

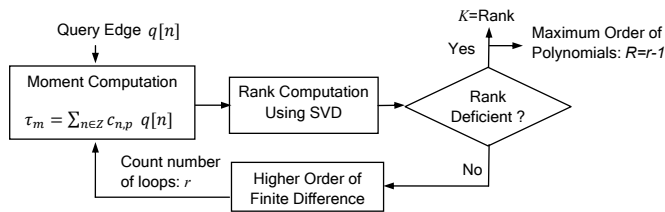


Fig. 4. Iterative algorithm for the retrieval of maximum order R of polynomials used in interpolation function $\lambda(t)$.

Next, the locations of all the discontinuities t_k and the continuous function $\hat{f}(t)$ can be retrieved using the annihilating filter method as discussed in [10]. Each t_k represents the location of samples $g[n]$ used in the interpolation. From the retrieval results in Figure 3(f), the distances between the differentiated Diracs $\hat{f}^{(2)}(t)$ are uniform and the sampling period T_1 can be estimated from the average of the distances.

The retrieved $\hat{f}(t)$ and T_1 then can be used to estimate the samples $g[n]$ through the reverse sampling. Finally, the retrieval of φ_1 can be further achieved using the best match between the samples and all possible dictionary elements as proposed in [6]. We omit this proof due to lack of space. Instead we focus on providing the sufficient conditions on φ_2 and T_2 that allow us to retrieve the chain.

Firstly, the maximum degree P of polynomial which the second kernel φ_2 can reproduce must be sufficiently large. From [10], the kernel must be able to reproduce polynomials of maximum degree $P \geq 2(R + 1)K - R - 2$ in order to achieve perfect reconstruction of a piecewise polynomial of maximum degree R with K discontinuities. In our case, the unit step input signal is sampled with uniform sampling period T_1 and the samples are then interpolated to the continuous domain again. The number of discontinuities can be computed as $K \leq \frac{L_1}{T_1} + 1$, where L_1 is the support of the first sampling kernel φ_1 . Therefore, the order P which provides the precisely retrieved results is given by $P \geq 2(R + 1)\frac{L_1}{T_1} + R$.

Secondly, we consider the role of t_2 which is now a constant with $t_2 > t_1$. The input, a rectangular pulse $f(t) = u(t - t_2) - u(t - t_1)$, is then acquired and reproduced by the chain. Since signal reconstruction creates a new group of K piecewise polynomials from samples of a unit step input, one needs to ensure that the two groups of piecewise polynomials are sufficiently distant in order to avoid overlap. The minimum interval required is greater than $2KT_1$. From [10], a piecewise polynomial function with two groups of K pieces of maximum degree R can influence an interval of size $2K(L_2 + R + 1)T_2$. One therefore can calculate the bound $T_1 > (L_2 + R + 1)T_2$, which imposes a constraint on the maximum sampling period T_2 . Here L_2 is the support of $\varphi_2(t)$.

When sampling signals with these requirements, a one-to-one mapping between discrete samples and chain structures is guaranteed. We conclude by providing a counter example to show that signals obtained from different acquisition chains cannot be distinguished when the sufficient conditions are violated. Let $q_a[n]$ and $q_b[n]$ be query discrete samples acquired from different chain structures. The signal $q_a[n]$ is obtained from a single acquisition of the step input $f_a(t) = u(t - (T_2 + \frac{T_2}{2}))$ using a box spline kernel [12] and $T_1 = T_2$ or $\varphi_{1a}(t) = \beta_0(\frac{t}{T_2})$. The signal, $q_b[n]$, is from a reacquisition. Given $f_b(t) = u(t - (T_2 + \frac{T_2}{4}))$ is the initial input, the signal is sampled using $\varphi_{1b}(t) = \beta_0(\frac{2t}{T_2})$ before the samples are linearly interpolated to $\hat{f}_b(t)$. From Figure 5, one can compute $q[n] = \langle f(t), \varphi_2(t/T_2 - n) \rangle$ and we have $q_a[n] = q_b[n] = [0 \ \frac{1}{2} \ 1 \ \frac{1}{2} \ 0]$. The signals from different chains become indistinguishable because the kernels used can reproduce polynomials to degree $P = 0$ which violates the condition. Thus, a one-to-one mapping is not guaranteed and the proposed algorithm cannot retrieve a unique chain solution.

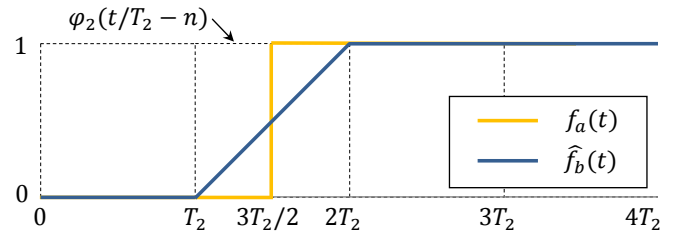


Fig. 5. Counter examples when sampling $f_a(t)$ and $\hat{f}_b(t)$ with a sampling kernel $\varphi(t) = \beta_0(\frac{t}{T_2})$

5. CONCLUSIONS

In this paper we have presented a theoretical scheme for the retrieval of a signal acquisition chain. With the theory of sampling signals with FRI, we are able to classify discrete samples from different stages in the chain model and to create a reacquisition detection algorithm. In addition, the method allows us to develop a retrieval algorithm for important parameters in the reacquisition chain. Lastly, a discussion on the conditions needed for the retrieval of the acquisition parameters was presented.

6. REFERENCES

- [1] A. Swaminathan, M. Wu, and K. J. Ray Liu, "A pattern classification framework for theoretical analysis of component forensics," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1665–1668, Apr. 2008.
- [2] A.C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of re-sampling," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 758–767, 2005.
- [3] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," *ACM Workshop on Multimedia and Security (MM&Sec)*, pp. 11–20, 2008.
- [4] N. Dalgaard, C. Mosquera, and F. Perez-Gonzalez, "On the role of differentiation for resampling detection," *IEEE International Conference on Image Processing*, 2010.
- [5] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 529–538, 2008.
- [6] T. Thongkamwitoon, H. Muammar, and P.L. Dragotti, "Identification of image acquisition chains using a dictionary of edge profiles," *EURASIP European Signal Processing Conference*, pp. 1757–1761, 2012.
- [7] C. E. Shannon, "Communication in the presence of noise," *Proceeding of IRE*, vol. 37, pp. 10–21, 1949.
- [8] M. Unser, "Sampling—50 Years after Shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, April 2000.
- [9] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, June 2002.
- [10] P.L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1741–1757, May 2007.
- [11] T. Blu, P. Thévenaz, and M. Unser, "MOMS: Maximal-order interpolation of minimal support," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1069–1080, July 2001.
- [12] M. Unser, A. Aldroubi, and M. Eden, "B-Spline signal processing: Part I—Theory," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 821–833, February 1993.