KCAD: Kinetic Cyber-Attack Detection Method for Cyber-Physical Additive Manufacturing Systems

Sujit Rokka Chhetri, Arquimedes Canedo†, Mohammad Abdullah Al Faruque
{schhetri, alfaruqu}@uci.edu
†arquimedes.canedo@siemens.com
Department of Electrical Engineering and Computer Science
University of California, Irvine, California, USA

ABSTRACT

Additive Manufacturing (AM) uses Cyber-Physical Systems (CPS) (e.g., 3D Printers) that are vulnerable to kinetic cyber-attacks. Kinetic cyber-attacks cause physical damage to the system from the cyber domain. In AM, kinetic cyber-attacks are realized by introducing flaws in the design of the 3D objects. These flaws may eventually compromise the structural integrity of the printed objects. In CPS, researchers have designed various attack detection method to detect the attacks on the integrity of the system. However, in AM, attack detection method is in its infancy. Moreover, analog emissions (such as acoustics, electromagnetic emissions, etc.) from the side-channels of AM have not been fully considered as a parameter for attack detection. To aid the security research in AM, this paper presents a novel attack detection method that is able to detect *zero-day kinetic cyber-attacks* on AM by identifying anomalous *analog emissions* which arise as an outcome of the attack. This is achieved by statistically estimating functions that map the relation between the *analog emissions* and the corresponding cyber domain data (such as G-code) to model the behavior of the system. Our method has been tested to detect potential *zero-day kinetic cyber-attacks* in fused deposition modeling based AM. These attacks can physically manifest to change various parameters of the 3D object, such as *speed*, *dimension*, and *movement axis*. Accuracy, defined as the capability of our method to detect the range of variations introduced to these parameters as a result of kinetic cyber-attacks, is 77.45%.

CCS Concepts

•Security and privacy → Intrusion detection systems; •Computer systems organization → Embedded systems;

Keywords

Intrusion Detection, Security, Cyber-Physical Systems, Additive Manufacturing, Kinetic Cyber-Attacks

ICCAD '16, November 07-10, 2016, Austin, TX, USA

 ISBN 978-1-4503-4466-1/16/11...$15.00
DOI: http://dx.doi.org/10.1145/2966986.2967050

1. INTRODUCTION & RELATED WORK

In Additive Manufacturing (AM), objects are created layer by layer [1]. Different types of AM vary according to the materials used and the technology incorporated for fusing the layers. Fused Deposition Modeling (FDM) is a common technology used in AM, where thermoplastic heated slightly above its melting point is deposited layer by layer to form a 3D object. It has been forecasted that 5.6 million 3D Printers will be shipped worldwide by 2019 [2]. High-end FDM based 3D Printers are already used in diverse fields such as automotive, aerospace, and medical devices [3]. For example, Airbus 350 is currently flying with more than 1000 3D Printed parts [4]. With its widespread in various fields, it is clear that existing security issues prevalent in Cyber-Physical Systems (CPS) will eventually affect AM as well. The major concern will arise due to kinetic cyber-attacks, which can cause physical damage, injury, or even death due to attacks originating in the cyber domain [5]. In CPS, effects of kinetic cyber-attacks have been recently highlighted by incidents such as the Stuxnet malware [6], Maroochy water breach [7], German steel mill cyber-attack [8], and security breaches in automobiles [9]. In AM, kinetic cyber-attack can find its way through the digital process chain to introduce various inconspicuous flaws in the 3D objects. If these objects are critical for the system, they can compromise the structural integrity and pose severe safety risk. For example, an inconspcuous void (less than 1 mm in dimension) placed in the 3D design of an American Society for Testing and Materials standard D638-10 tensile test specimen, reduced its mechanical strength to carry load by 14% [10].

The security concerns in CPS are not new [11, 12], and in fact various attack detection methods have been designed for the identification and detection of attacks [13]. However, AM has not received much attention for the research in attack detection methods. In [10], authors present the potential attack vectors for AM digital process chain and recommend securing the process chain by incorporating software checks, hashing, and process monitoring through side-channel. In [14], authors present signature based attack detection method leveraging the concept of integrated circuit Trojan detection from side-channel analysis and system health monitoring. However, it should be noted that signature based technique requires acquiring the signature of a baseline structure every time it is created, which is counter-intuitive for rapid prototyping capability of AM. In [15], authors list the threat surface and describe its effects on the manufacturing parameters. In AM, any variation made in

the design of the 3D object is manifested physically. Moreover, during the printing stage, machine itself unintentionally emits *analog emissions* from the side-channels (such as acoustics, electromagnetic radiation, power, etc.). Thus, we propose a novel Kinetic Cyber-Attack Detection (KCAD) method that uses statistical modeling of the AM system to detect the anomalous *analog emission* which can arise as a result of potential *zero-day kinetic cyber-attacks*.

1.1 Motivation

The fundamental motivation for KCAD method comes from the fact that in CPS, the information flow in the cyber domain has at least one corresponding signal flow in the physical domain [16]. These signals in the physical domain actuate the physical processes, and this actuation converts energy from one form to the another. This phenomenon allows us to monitor the unintentionaly leaked *analog emissions* which have high *mutual information* with the corresponding control signals. *Analog emissions* have been used in system health monitoring and prognostics to infer about the current state of the system, and also for quality control in manufacturing [17, 18]. However, traditional quality control system focus in measuring the key quality characteristics, and kinetic cyber-attack on various features may not be detected by such systems [14]. On the other hand these *analog emissions* are also used to breach the confidentiality of the system [19]. However, incorporating statistical method, the acquired *analog emissions* from the side-channel corresponding to the *control signals* can be used to model the behavior of the system. And this model may be used for detecting the intrusion in the system [20].

1.2 Problem and Research Challenges

The current challenge for securing the CPS is understanding its unique properties and vulnerabilities, and preventing the possibility of an attack [21]. However, *zero-day vulnerabilities* of the system are hard to detect during design time. In such scenarios, detection of such attacks is the best possible defense. After detection, we can take measures such as halting the system, perform corrective actions, etc., to mitigate the effects of the attack. Details of the mitigation methods are out of the scope of this paper. In AM, the problem for designing a *zero-day kinetic cyber-attack* detection method poses the following key challenges:

1. Detecting the intrusion/attack that can occur at various points of the digital process chain of AM, and affect the dynamics of the system.
2. Making it non-intrusive so that it can be used in legacy AM systems.
3. Complementing the detection method with the physical and process knowledge from AM.

1.3 Our Novel Contributions

To address the above mentioned challenges, we propose a novel attack detection method for detecting the effects of possible *kinetic attacks* in the digital process chain of the additive manufacturing that employs:

1. **Modeling of an Adversary (Section 3)** to understand the various attack points in the digital process chain for effective implementation of the attack detection method.
2. **Statistical Estimation (Section 4.2)** to model the behavior of the AM system by analyzing the relationship between the *analog emissions* and the control signals.
3. **Analysis of Analog Emission (Section 4.3)** to use it as a parameter from the side-channel for estimating the relationship between the *analog emissions* and the control parameters using **mutual information** as the relation measurement metric **(Section 4.1)**.

2. BACKGROUND

In FDM based 3D Printers, the digital process chain consists of various steps [1]. The first step is to design the 3D object using Computer Aided Design (CAD) tools. This CAD model is then converted to STereoLithograhy (STL) format which uses series of triangles to model the surface geometry. Conversion to STL format is automatically done by the CAD software. Slicing algorithms are used to convert the STL file into G/M-code which consists of layer by layer description of the 3D model. G/M-code is machine specific and the AM machine's firmware can convert it to corresponding control signals to actuate the various physical components. G-codes are responsible for handling the motion while printing. It is responsible for determining the speed of printing along different axis, as well as the extrusion amount to be deposited in each printing step. As an example, $G1\ F2100\ X5\ Y6\ Z1.2\ E2.1$ represents a single line of G-code for controlling the movement of the nozzle, where $G1$ means coordinated linear motion, F defines travel feed rate (speed) which is measured in mm/min, and distance and extrusion are measured in mm. The M-codes are used for controlling the machine settings such as temperature, coolant, etc.

3. ADVERSARY MODEL

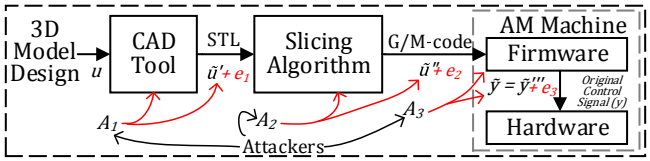


Figure 1: Adversary Attack Points.

In AM, information (3D design specification) flows through the digital process chain, and is finally converted as the control signals in the machine. In the adversary model, an attacker can infiltrate at various stages of the process chain (see Figure 1) to alter the ***integrity*** of the tools, algorithms, and firmware. Moreover, they may add exogenous inputs e_1, e_2 and e_3 during the transfer of information from one stage to the other in the digital process chain. Let us consider y to be the control signals to the physical components of the AM machine, provided u is the true information. In this case, $\tilde{u}'$, $\tilde{u}''$ and $\tilde{y}'''$ are the false information produced in the process chain due to the compromised ***integrity*** of the CAD tool, slicing algorithm, and the firmware respectively. We can observe that the attacks on the digital process chain by different attackers A_1, A_2 and A_3 always results in the modification of the control signals y. This phenomenon is what separates the CPS from traditional information and technology systems. In FDM based AM, the outcome of kinetic attack will result in the variation of control signal y to $\tilde{y}$, such that it modifies the initial 3D design of the object.

REMARK 1. *In FDM based AM, the change in the information flow by an attacker will result in the change of control parameters $y = [v, a, t, d]$ responsible for controlling the dynamics of the machine.*

Where $v = [v_x, v_y, v_z, v_e]$ represents the speed of the nozzle in different axis, along with the speed of extrusion, and $v_{i\in\{x,y,z,e\}} \mathbb{R}_{\geqslant 0}$. $a = [a_x, a_y, a_z, a_{xy}, a_{xz}, a_{yz}, a_{xyz}]$ represents the axis of movement such that $a_{i\in\{x,y,z,xy,xz,yz,xyz\}} \in \{0, 1\}$, and $a = 1$ represents presence of movement in the given axis and $a = 0$ represents absence of movement. $t \in \mathbb{R}_{\geqslant 0}$ represents the temperature of the nozzle. $d = [d_x, d_y, d_z, d_e]$ represents the distance of the nozzle in different axis, along with the amount of extrusion, and $d_{i\in\{x,y,z,e\}} \in \mathbb{R}$. Hence, the kinetic cyber-attack in FDM based AM, will cause the information u to be altered such that the final control parameters to the physical components are altered to $\tilde{y} = [\tilde{v}, \tilde{a}, \tilde{t}, \tilde{d}]$.

DEFINITION 1. *With reference to remark 1, the kinetic cyber-attack in FDM can be defined as change in the information flow u in the digital process chain such that:*

$$\begin{bmatrix} v \\ a \\ d \\ t \end{bmatrix} \pm \begin{bmatrix} e_v \\ e_a \\ e_d \\ e_t \end{bmatrix} = \begin{bmatrix} \tilde{v} \\ \tilde{a} \\ \tilde{d} \\ \tilde{t} \end{bmatrix} \tag{1}$$

and $\tilde{y} \neq y$, when $\sum_{i\in\{v,a,d,t\}} e_i > 0$. Here $\{e_v, e_d, e_t\} \in \mathbb{R}_{\geqslant 0}$, $e_a \in \{0, 1\}$.

4. KCAD METHOD

Let $Y \rightarrow O$ be a side-channel, where Y and O represent random variables denoting control information parameters and observed *analog emission* respectively. Then, KCAD method leverages the fact that these variables have high mutual information.

4.1 Mutual Information

REMARK 2. *Let the observed analog emissions be $o(t)$, then the control parameters, $y = [v, a, t, d]$, responsible for controlling the dynamics of the system, emit analog emissions such that the mutual information $\{I(V;O), I(A;O), I(T;O), I(D;O)\} > 0$. Where (V, A, T, D) are random variables.*

The random variables O, V, T, and D are continuous where as A is discrete. Let $f(o)$, $f(v)$, $f(t)$, and $f(d)$ be probability density function (*pdf*) of continuous random variables O, V, T, and D respectively. And for all $k, j \in \{o, v, t, d\}$, let $f_{k\neq j}(k, j) = f_{k\neq j}(j, k)$ be the joint *pdf*. The conditional *pdf* for these random variables is then defined as $f_{k\neq j}(k|j) = \frac{f_{k\neq j}(k,j)}{f(j)}$. Then, we can calculate the differential entropy of these random variables as follows:

$$h_{K\in\{O,V,T,D\}}(K) = -\int f(k) log(f(k)) dk \tag{2}$$

Similarly, the conditional differential entropy of these random variables can be given as follows:

$$h_{K,J\in\{O,V,T,D\}}(K|J)_{K\neq J} = -\int\int f(k, j) log(f(k|j)) dk dj \tag{3}$$

The mutual information between the observed *analog emission* and the control parameters can be given as follows:

$$I_{K\in\{V,T,D\}}(K; O) = h(K) - h(K|O) \tag{4}$$

Let $f(a)$ be a probability distribution function of the discrete random variable A. Then the entropy of A can be calculated as follows:

$$H(A) = -\sum f(a) log(f(a)) \tag{5}$$

For calculating the mutual information between O and A, we can divide the values of O into bins of length ε. If $H(O_\varepsilon)$ be the entropy of O after discretization, we have $h(O) = \lim_{\varepsilon \rightarrow 0} [H(O_\varepsilon) + log(\varepsilon)]$. And the mutual information can be calculated as:

$$I(A; O) = H(A) - H(A|O_\varepsilon) \tag{6}$$

The calculation of mutual information between two continuous random variables requires estimation of the probability density functions, which are then used in Equations 2 and 3. Kernel probability density estimation can be used for estimating the *pdf* based on the experimental data as:

$$\tilde{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - x_i}{h}) \tag{7}$$

Where K is a real-valued integrable kernel function and h the bandwidth. There are various kernel function that can be used for the estimation.

Our proposed KCAD method can be defined as a pipeline of deterministic algorithms which takes continuous observable *analog emissions* $o(t)$ as input along with the information flow U in the form of G-codes $G_t = [g_1, g_2, g_3, ..., g_t]$ from which the control parameters v, a, d, and t are parsed. The information (U) acquired by KCAD is assumed to be from a secure channel and free from any modification. KCAD method infers the *analog emission* O based on the given control parameters. Given the presence of a kinetic cyber-attack, the control parameters are changed to $\tilde{y} = [\tilde{v}, \tilde{a}, \tilde{d}, \tilde{t}]$ with observed *analog emissions* being $\tilde{O}$, such that $|\tilde{O} - O| = e$. And for $e > e^T$, emission variation threshold, the output of the detection system is *True*, denoting presence of an attack.

DEFINITION 2. *Attack Detectability: Given the input O and G with parsed variables v, a, d, and t, kinetic attack to the system, such that $\sum_{i\in\{v,a,d,t\}} e_i > 0$, is detected by KCAD method if its output is true.*

4.2 KCAD Architecture

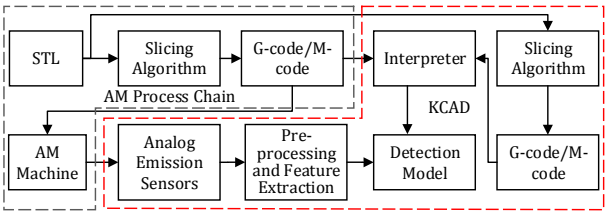


Figure 2: Architecture of KCAD Method.

The architecture of the simplified KCAD method is shown in Figure 2. One of the advantages of this method is that it can be placed to monitor the information flow in any of the stages of the digital process chain. With this, the attack on the integrity of any tools, firmware, and algorithms can be detected if it has corresponding affect on the dynamics of the system. KCAD runs in parallel to the AM while it is printing, and non-intrusively and continuously acquires the observable *analog emissions*. The components of KCAD that work in parallel to the AM, depends on which point of the process chain we feed the information to it. For example, if we want to detect the attack on the integrity of the

firmware of the AM, the input to the KCAD can just be the G-code/M-code. The slicing algorithm in it can thus be switched off. However, the channel through which the information passes to the KCAD method from different point of the digital process chain is assumed to be secure. This means that the KCAD method will always receive original/unmodified cyber data from the digital process chain.

Analog Emission Sensors: Various sensors (piezoelectric, current, electromagnetic, etc.) can be used to monitor the *analog emissions* from the AM system. Given the integrity attack that introduces values e_v, e_a, e_d, and e_t, in the control parameters (v, a, d, t), the sampling frequency (F_s) and bandwidth (B) should be such that it can measure corresponding changes e_v, e_a, e_d, and e_t in the *analog emissions* $o(t)$. Moreover, distance and angle of placement of the sensors also affect the Signal to Noise Ratio (SNR) given as:

$$SNR_{dB} = 10log_{10}(\frac{P_{Signal}}{P_{Noise}}) \tag{8}$$

Hence, the choice and placement of the sensors depends on the choice of side-channel and the relation between the *analog emissions* and the control parameters. For choice of side-channel and the corresponding *analog emissions*, Equations 4 and 6 can be used as a measure of relation between the observable *analog emissions* and the control parameters. Based on this measurement, the observed values can either be incorporated or discarded from group of features to be used for estimating the behavior of the system in KCAD.

Pre-processing and Feature Extraction: Pre-processing is done to improve the SNR by removing the known noise signals from the *analog emissions* that are independent of the control parameters. If observable *analog emission*, $o(t) \neq f(y(t))$, where y represents the control parameters, then the observed *analog emission* can be considered as noise. The signals acquired by sensors can be too large for the estimation algorithms used in the detection model. Hence, it is necessary to extract only the informative values from the signal to improve the processing time of the detection model. For each observed signal various values (features) are derived using the original signal.

$$O_t = \begin{bmatrix} o_1^{f^1} & o_1^{f^2} & o_1^{f^3} & \dots & o_1^{f^n} \\ o_2^{f^1} & o_2^{f^2} & o_2^{f^3} & \dots & o_2^{f^n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ o_t^{f^1} & o_t^{f^2} & o_t^{f^3} & \dots & o_t^{f^n} \end{bmatrix} \tag{9}$$

Where the column represents the number of features and the row represents the discretized values of the signal $o(t)$. The type of features extracted is specific to the observed *analog emissions* (for acoustics see Section 4.2). For reducing the dimension of the extracted features, principle component analysis is used.

Interpreter: Each block (line) of instruction (G/M-code) sent to the AM consists of control parameters. These instructions or numerical control codes are converted to the canonical machining commands using interpreters such as NIST RS274NGC. In our KCAD method, a lighter version of arduino G-code and NIST RS274NGC Interpreter is used to extract the control signals v, a, d, and t. These signals are then sent to the detection model.

Detection Model: The detection model uses the supervised learning approach to estimate the function $\hat{f}_i(O_t, \alpha_n)$ with $i = 1, 2, 3, 4$ using the training data-set of observed *analog emissions*. For each control parameters we estimate the function with respective parameter α. Various predictive models can be estimated based on the initial training data-sets. The training data-sets will require to balance the trade-off between bias and variance to find optimal parameter α. $k-$fold cross-validation is used to perform data driven validation for the estimated function $\hat{f}_n(.)$. For the regression function estimation, learning algorithms such as *Gradient Boosting Regressor (GBR)*, *Ridge Regression*, *Stochastic Gradient Descent Regression (SGDR)*, *Bayesian Ridge Regression (BRidge)*, *Passive Aggressive Regression (PAR)*, *Decision Tree Regression (DTR)*, *Elastic Net Regression (ENet)*, *Linear Regression with Lasso*, and *k-Nearest Neighbor Regression (kNN)* is used. Each of these models is compared using metrics such as *explained variance*, *Mean Absolute Error (MAE)*, *Mean squared error (MSE)*, *Median Absolute Error*, and R^2 *Score*. For function that needs to be estimated for classification, classifiers such as *Support Vector Machine (SVC)* with *Linear* and *Radial Basis Function (RBF)* as kernels, *Logistic Regression*, *Stochastic Gradient Descent Classifiers*, *Ensemble of AdaBoost* and *Gradient Boosting* are used. They are compared using Receiver Operating Characteristic (ROC) curves, and the model with the least error is selected.

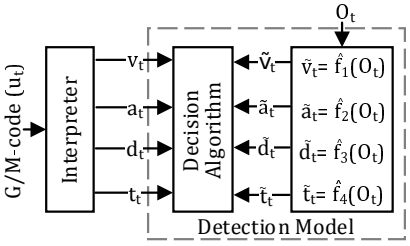


Figure 3: Detection Model.

Detection Algorithm: The detection algorithm compares the real control parameters given by the interpreter and the values calculated by the estimated functions $\hat{f}_n(.)$.

Algorithm 1: Detection Algorithm.

Input: Real and Estimated Control Parameters $[v, a, d, t], [\tilde{v}, \tilde{a}, \tilde{d}, \tilde{t}]$
Output: Attack Flag F_A

1 Define Error Thresholds $e_v^T, e_a^T, e_d^T, e_t^T$
2 Initialize $f_v = 0, f_a = 0, f_d = 0, f_t = 0$
3 **for** *each* $i \in v, a, d, t$ **do**
4 **if** $|i - \tilde{i}| > e_i^T$ **then**
5 $f_i = 1$
6 **if** $\sum_{i \in \{v,a,d,t\}} f_i \geq 1$ **then**
7 $F_A = 1$
8 **return** F_A

In Algorithm 1, in line 1, the error variation thresholds are defined. This value is based on the accuracy of the estimated functions $\hat{f}_n(.)$ during the training phase of the detection model. Lines 3 to 5 determines if the estimated control parameters and the real control parameters vary more than the error threshold. Lines 6-7 set the attack detection flag to high if any of the control parameters varies more than the error variation threshold. Finally in line 8 the attack flag is returned.

Offline vs Online Model Estimation: The model function $\hat{f}_n$ estimation can be done either online or offline. This is necessary to consider because of the fact that AM machine will have varying observable *analog emission* over long period of time due to wear and tear of the mechanical structures. Offline function estimation can have shorter response time. However, online estimation can be done for higher accuracy with longer response time.

4.3 Acoustic Analog Emissions

Acoustic *analog emissions* is one of the observable emissions in cyber-physical additive manufacturing system. The fundamental working principle behind KCAD method is that the *analog emissions* $o(t)$ must have high mutual information with the control parameters $y(t)$. As a proof of concept of the KCAD method, we will use acoustics as the observable *analog emissions* to detect the presence of kinetic cyber-attack on AM. However, it is trivial to infer that the observed *analog emission* will have weak relation with the control parameter temperature t. Therefore, only a kinetic cyber-attack affecting the control parameters v, a, and d will be considered. The main source of acoustics in FDM based 3D Printers are the vibration of the stepper motors. These 3D Printers consists of at least one stepper motor to control the movement of the nozzle of the printer in each axis (x, y, and z axis) [22]. These stepper motor consists of rotor (permanent magnet) and the stator (electromagnet). The varying radial electromagnetic force acting on the stator of the stepper motor produces vibration [23–25]. This vibration is the source of acoustics. The radial electromagnetic force is controlled by the control parameters, such as *speed* of the movement of the motor. However, the natural frequency of the stator is determined by the load, connected frame, and the structure of the stator [26]. Hence, resonance occurs when the vibration produced by the radial electromagnetic matches the harmonics of the natural frequency of the stator. This resonance frequency is different for different stepper motors responsible for moving the 3D Printer's nozzle in x, y, and z. This will allow us to estimate functions to separate the movement in different axes. The *analog emission* sensors for capturing the acoustic emissions will require sampling frequency of more than 40 kHz to capture the range of audible sound 20 Hz to 20 kHz. During the pre-processing stage digital filter is used to remove, low and high frequency noise.

Algorithm 2: Dynamic Window Size Determination.

Input: Observed Analog Emission $o(t)$
Output: Dynamic Windows $w = [w_1, w_2, \ldots, w_n]$
1 Initialize $n = 1, i_{previous} = 1$
2 Extract Features O_t // $n \rightarrow$`Number of Features`
3 **for** $i = 2$ *to* t **do**
4 **if** $|\sqrt{(o_i^{f^1} - o_{i-1}^{f^1})^2 + \ldots + (o_i^{f^n} - o_{i-1}^{f^n})^2}| > d^T$ **then** // $d^T \rightarrow$`Threshold Distance`
5 $w_n = i - i_{previous} + 1$
6 $i_{previous} = i + 1$
7 $n = n + 1$
8 **return** w

Dynamic time warping is used to dynamically assign the window size w for feature extraction. However, an initial fixed length window size (10-50 ms) will be used to extract features size as *Zero Crossing Rate*, *Energy Entropy*, *Spectral Entropy*, and *Mel Frequency Cepstral Coefficients (MFCCs)*. Using these features, euclidean distance is measured to define the dynamic window size for accurate feature extraction. As shown in Algorithm 2, line 4 measures the euclidean distance between the features of the previous analog observation with the current *analog emission* in discrete time series (based on the sampling frequency). If this distance is greater than the threshold distance d^T, which is determined by measuring the distance between the *analog emission* of training data-sets. As windowing is done to extract the features from the observed *analog emissions* $o(t)$, we can determine the control parameter v from d and vice-versa, where $v = d/w$. Where w is length of the window in seconds.

REMARK 3. *For FDM based AM, KCAD method will be able to monitor any kinetic attacks modifying the control parameters v, a, and d by analyzing the variation in the acoustic analog emissions to the corresponding control parameters v and a.*

4.4 Performance Metrics

The performance of an attack detection method can be measured with two metrics *True Positive Rate (TPR)* and *True Negative Rate (TNR)*.

$$TPR = \frac{TP}{TP + FN} \tag{10}$$

where $True\ Positive\ (TP)$ is the total number of positive detection when there is an attack in the system, and $False\ Negative\ (FN)$ is the total number of negative detection during the presence of an attack. Similarly,

$$TNR = \frac{TN}{TN + FP} \tag{11}$$

where $True\ Negative\ (TN)$ is the total number of positive detection when there isn't any attack to the system, and $False\ Positive\ (FP)$ is the total number of positive detection when there isn't any attack to the system. Then, the accuracy of the system can be measured as follows:

$$Accuracy = \frac{TP + TN}{Total\ Sample} \tag{12}$$

5. EXPERIMENTAL RESULTS

5.1 Experimental Setup

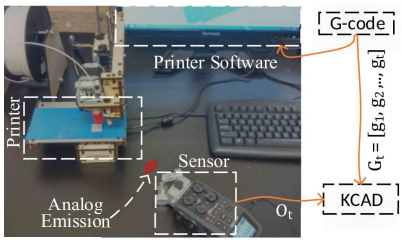


Figure 4: KCAD method Experimental Setup.

The experimental setup for the KCAD method evaluation is shown in Figure 4. This evaluation tests KCAD performance against integrity attacks on the printer's firmware. In a simple scenario, a modified firmware is installed in the 3D Printer by an attacker. This attack modifies the control signal to the 3D Printer, which introduces variation in the geometry of the 3D object. The audio recorder is placed at an optimal distance to acquire the acoustics. In the experiment, the recorder is placed at 45° angle to the x and y axis to acquire the variation of the signal in both

Table 1: Mutual Information in Bits Between Features and Control Parameters.

Parameters/Features	$o_t^{f^1}$	$o_t^{f^2}$	$o_t^{f^3}$	$o_t^{f^4}$	$o_t^{f^5}$	$o_t^{f^6}$	$o_t^{f^7}$	$o_t^{f^8}$	$o_t^{f^9}$	$o_t^{f^{10}}$	$o_t^{f^{11}}$	$o_t^{f^{12}}$	$o_t^{f^{13}}$	$o_t^{f^{14}}$
v_x	0.79	0.95	0.16	0.81	0.74	0.76	0.35	1.18	1.51	0.31	0.55	0.69	0.75	0.71
v_y	0.27	0.67	0.11	0.43	0.41	0.24	0.23	1.31	1.10	0.23	0.44	0.45	0.55	0.60
v_z	0.16	0.07	0.0806	0.08	0.07	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.09
a	0.58	0.82	0.18	0.69	0.69	0.53	0.44	1.91	1.17	0.64	0.74	0.22	0.40	0.39

direction with a single recorder. During the training phase, G-codes are written to move the nozzle of the printer in various x and y axis directions with various printing speeds (400 mm/min to 4500 mm/min with 100 mm/min step size). These speed ranges are machine specific. Using this training data, we estimated the model function for the control parameters v_x, v_y, and $a = [a_x, a_y, a_z, a_{xy}]$. Using a similar approach, model functions for parameters v_e, v_z, different directions, and $a = [a_{yz}, a_{xz}, a_{xyz}]$ can be estimated.

5.2 Mutual Information Calculation

To demonstrate the dependency of the acoustic *analog emissions* with the control parameters $v = [v_x, v_y, v_z]$ and, the $a = [a_x, a_y, a_z, a_{xy}]$ mutual information between them is calculated. For control parameters a, it is treated as a discrete random variable with different labels depending on the combination of axis movements. Table 1 shows the mutual information calculated for the control parameters and the features extracted from the observed *analog emissions*. We can see that different features have varying mutual information with the control parameters. Also, it can be observed that the mutual information between the speed in $z-axis$ and the *analog emission* is comparatively low. This is due to the reason that speed in $z-axis$ is almost constant in most of the 3D Printers. The estimation function utilizes different features on the basis of the mutual information and principle component analysis to select the features that are most relevant.

Table 2: Regression Models Comaprison for $\hat{f}_v(.)$.

Model	MSE	Variance	MAE	Median AE	R^2
GBR	0.0076	0.9923	0.0037	0.0167	0.9923
Ridge	0.0147	0.9854	0.0090	0.0775	0.9851
SGDR	0.0148	0.9852	0.0090	0.0785	0.9850
BRidge	0.0149	0.9852	0.0089	0.0772	0.9849
PAR	0.0183	0.9817	0.0095	0.0899	0.9815
DTR	0.0258	0.9741	0.0090	0.0582	0.9740
ENet	0.0786	0.9212	0.0210	0.1990	0.9208
Lasso	0.1015	0.8980	0.0241	0.2331	0.8976
kNN	0.0025	0.4997	0.0014	0.0182	0.4997

5.3 Model Function Estimation

The function, $\hat{f}_i(O_t, \alpha_n)$, estimation is the fundamental step in our KCAD method. The parameter $[\alpha_1, \alpha_2, \ldots, \alpha_n]$ are responsible for minimizing the cost functions used by the learning algorithms. This estimation is done in the training phase. Based on the relation between the control parameters and the features extracted from observed *analog emissions*, the estimated functions can be used for regression or classification. The relation between the control parameter v and observed *analog emissions* $o(t)$ both being continuous random variable can be estimated using regression algorithms. Where as, the parameter a is a discrete random variables and and we have to use classifiers for estimating the function $\hat{f}(.)$. We will use various algorithms and perform the comparison between them to select the function $\hat{f}(.)$, that gives the least error.

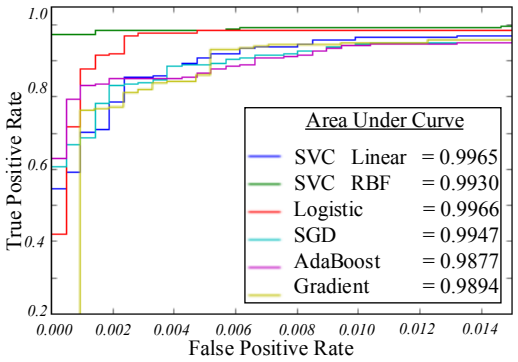


Figure 5: ROC of Classifiers.

Regression: Function $\hat{f}_{[v_x,v_y]}(O_t, \alpha_n)$ is estimated for the control parameter v based on O_t. From Table 2, it is clear that *Gradient Boosting Regression*, outperforms rest of the regression models in terms of the error metrics. Hence, it is selected to estimate the function for the control parameters (v_x, v_y). Function for v_z is not estimated as speed in z-axis is constant.

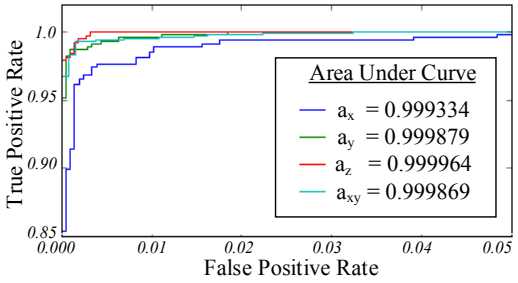


Figure 6: ROC of Multi-Classes, One versus Rest.

Classification: For the parameter a, various functions estimated along with their performance is compared in Figure 5. The estimation function is treated as one versus the rest classification and the ROC is calculated as an average of ROC of all the classes i.e various movement axis, $[a_x, a_y, a_z, a_{xy}]$. From Figure 5, *logistic regression* classifier is selected as the optimal function for the classification. In Figure 6, the ROC curve of various classes calculated as a one versus rest classification using *Logistic Regression* classifier is presented. Since, the area under the curve for the $z-axis$ movement class is higher, the detection of presence of nozzle movement in $z-axis$ is easiest compared to the other classes. With these functions estimated for various control parameters, KCAD can effectively detect the variation in the *analog emissions* with the expected emissions in the presence of kinetic cyber-attacks on the firmware of the 3D Printer.

5.4 Results for Detection of Kinetic Attack

We developed a zero-day kinetic attack on the 3D Printer's firmware to test the KCAD method. Our attack modifies the

nozzle speed in the x and y direction while printing, thus effectively changing the dimension of the object. Additionally, the axis values are changed resulting in the deformation of the object. For detecting the variation of speed on $x-axis$ and $y-axis$, the speed is varied from 600 mm/min to 3500 mm/min, while the attack is assumed to introduce range of variation in the original speed i.e. from 50 mm/min to 1000 mm/min. From the function estimation, error threshold for speed e_v^T is set as 200 mm/min.

Table 3: True Positives for Speed Variation.

δv	TP for Speed (*100 mm/min)									Total
mm/min	7	9	12	15	17	20	25	30	35	TP
1000	16	16	16	16	16	16	16	16	16	144
500	16	16	16	16	16	14	15	16	14	139
300	16	16	16	13	10	7	7	8	7	92
200	13	16	10	9	8	6	8	8	6	84
TPR										0.7968

Table 4: False Positives for Speed Variation.

δv	FP for Speed (*100 mm/min)									Total
mm/min	7	9	12	15	17	20	25	30	35	FP
1000	5	4	3	6	5	7	7	6	8	51
500	4	3	2	5	5	8	6	8	8	54
300	2	2	3	5	5	7	6	6	7	47
200	3	1	3	5	2	8	7	8	7	44
FPR										0.3402

Table 3 and 4 show that for higher speed and lower speed variation (δv), the true positives are lower compared to the low speed and high speed variation. However, the false positives are higher for higher speeds.

Table 5: True Positives for Distance Variation.

δd	TP for Speed (*100 mm/min)									Total
mm	7	9	12	15	17	20	25	30	35	TP
20	16	16	16	16	16	16	16	14	16	142
10	16	15	16	16	16	15	14	13	14	135
5	16	13	14	10	12	14	13	12	15	119
3	14	10	12	10	11	10	12	11	13	103
TPR										0.8663

Table 6: False Positives for Distance Variation.

δd	FP for Speed (*100 mm/min)									Total
mm	7	9	12	15	17	20	25	30	35	FP
20	4	4	4	6	4	3	3	5	4	37
10	3	4	4	5	4	4	4	5	5	38
5	2	5	5	5	5	4	5	4	7	42
3	2	4	7	7	5	2	5	5	8	45
FPR										0.2812

Intuitively, this is due to the fact that feature extracted prominently consists of MFCC, which focuses on extracting more features from lower frequency range rather than higher frequencies causing poor function estimation for higher speeds. KCAD accuracy for detection of attack on control parameter v can be calculated using Equation 12 as 72.83%. For testing the performance of KCAD method for detecting the change in the control parameter d introduced by the modified firmware, distance is varied from 3 mm to 20 mm. The error threshold for the distance is selected ad 3 mm based on the error in the estimated function.
Table 5 shows that the number of true positive is decreasing with the increasing speed and lower distance variation δd. Moreover, the number of false positive is increasing with the increasing speed. Using Equation 12, the accuracy for detection of attack on control parameter d is calculated to be 79.25%.

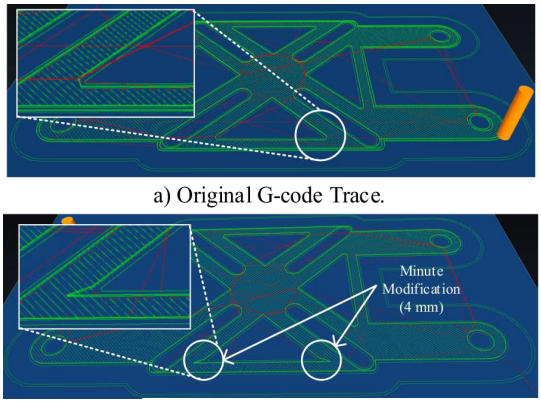
Table 7: True Positives for Axis Variation.

$Axis, Total$	TP for Speed (*100 mm/min)									Total
TP	7	9	12	15	17	20	25	30	35	TP
$a_x, 32$	32	32	28	28	24	21	21	20	19	225
$a_y, 32$	31	32	27	25	23	19	19	19	18	213
$a_{xy}, 24$	20	21	21	19	19	20	19	18	17	174
$a_z, 24$	24	20	19	18	18	19	18	19	18	173
TPR										0.7787

Table 8: False Positives for Axis Variation.

$Axis, Total$	FP for Speed (*100 mm/min)									Total
TN	7	9	12	15	17	20	25	30	35	FP
$a_x, 32$	2	2	2	3	4	6	6	7	9	41
$a_y, 32$	2	3	4	2	4	5	4	9	8	41
$a_{xy}, 24$	4	4	6	6	7	5	6	7	8	53
$a_z, 24$	5	6	7	8	7	8	7	8	8	64
FPR										0.1974

We modified the movement in x, y, z, and xy axis to measure the KCAD performance against firmware modification attacks that vary the control parameter a, which determine the movement axis. Table 7 shows that the true positive rate is decreasing with increase in speed and false positive rate in increasing with increase in the speed. The accuracy for detection of attack on control parameter a is 79.07%. Our KCAD method relies on the fact that any *zero-day kinetic cyber-attack* results in variation of control parameters v, a, and d. Hence, treating these parameters as synthetic benchmarks, the accuracy of the KCAD method, for measuring the variation of various control parameters v, a, and d is 77.45%.



a) Original G-code Trace.

b) G-code Trace after Kinetic Attack.

Figure 7: Attack on Base Plate of a Quad Copter.

5.5 Test Case: Base Plate of a Quad Copter

As a test case, we present an analysis on a flight controller base plate [27], which is a part of quadcopter that can be printed using a 3D Printer. We assume that the modified firmware, as a result of *zero-day kinetic cyber-attack*, introduces variation in the certain part of the code by adding 4 mm to the $x-axis$ movement distance. Such small changes in the design of an object can compromise their structural integrity during use, and lead to a catastrophic failure. The original design of the base plate is shown in Figure 7 (a). As a result of an attack, minute modification introduced in the design is shown in Figure 7 (b). This modification might not be visible to human eyes, however it compromises the structural integrity of the base plate. KCAD

method detected the variation in the x-axis introduced in all three layers by the firmware modification attack.

6. LIMITATION AND FUTURE WORK

In this section, we provide an open discussion on the limitations of our KCAD method:

1. Printer Variation: KCAD is machine specific. Hence, for implementation, the function estimation (training) has to be conducted before it can be implemented. Different FDM based 3D Printers will emit different acoustic signature based on the type of motors used and the structure of the frame. This has to be studied before KCAD can be implemented. Different *analog emissions* and their respective features have to be analyzed for model function estimation. However, this has to be done only once before the implementation.

2. Complex Attacks: In the experimental section, we have tested the KCAD with firmware modification attacks introducing simple variation in the x, y axis. However, more complex variation can be introduced by the attack, with attacks modifying both the x and y axis together. In such scenarios, more function have to be estimated for combination of control parameters.

3. Sensor Placement: The *analog emissions* sensors placement should not obstruct the printing process. However, it must be able to capture the emissions with high SNR. This requires analyzing various sensors and its optimal position. It might have to be placed inside the system for better SNR. However, this can only be done if it does not obstruct the printing process.

7. CONCLUSIONS

This paper presented a novel kinetic cyber-attack detection method, which can be placed non-intrusively, and can monitor the system during run-time. We have performed the analysis of acoustic *analog emissions* to measure the feasibility of such system, and demonstrated that acoustics have high mutual information with the control parameters parsed from the cyber domain data. We have tested our system on a FDM based 3D Printer, assuming that integrity attack on the printer firmware eventually modifies the control parameters to the physical components. We have tested the performance of our method with variation of speed, distance, and axis that can be an outcome of the kinetic cyber attack in the digital process chain of the additive manufacturing. KCAD method achieves a high accuracy of 77.45%. This provides a good starting point and proof of concept for the proposed attack detection method.

8. REFERENCES

[1] I. Gibson, D. W. Rosen, B. Stucker, *et al.*, *Additive manufacturing technologies*. Springer, 2010.

[2] "Forecast: 3D Printers, Worldwide, 2015.," *www.gartner.com*, 2015.

[3] H. Lipson, "3D Printing, Now and Beyond," *Stratasys, www.stratasys.com*, 2015.

[4] B. Krassenstein, "New Airbus A350 XWB aircraft contains over 1,000 3D printed parts." 3Dprint, https://3dprint.com, 2015.

[5] S. D. Applegate, "The dawn of kinetic cyber," in *Cyber Conflict (CyCon), 2013 5th International Conference on*, pp. 1–15, IEEE, 2013.

[6] N. Falliere *et al.*, "W32. stuxnet dossier," *White paper, Symantec Corp., Security Response*, vol. 5, 2011.

[7] J. Slay and M. Miller, *Lessons learned from the maroochy water breach*. Springer, 2007.

[8] R. M. Lee, M. J. Assante, and T. Conway, "German steel mill cyber attack," *Industrial Control Systems*, vol. 30, 2014.

[9] K. Koscher *et al.*, "Experimental security analysis of a modern automobile," in *Security and Privacy (SP), 2010 IEEE Symposium on*, pp. 447–462, IEEE, 2010.

[10] L. Sturm *et al.*, "Cyber-physical vunerabilities in additive manufacturing systems," *Context*, 2014.

[11] M. Al Faruque *et al.*, "Design methodologies for securing cyber-physical systems," in *Proceedings of the 10th International Conference on Hardware/Software Codesign and System Synthesis*, 2015.

[12] J. Wan, A. B. Lopez, and M. A. Al Faruque, "Exploiting wireless channel randomness to generate keys for automotive cyber-physical system security," in *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)*, pp. 1–10, IEEE, 2016.

[13] R. Mitchell and I.-R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Computing Surveys (CSUR)*, 2014.

[14] H. Vincent *et al.*, "Trojan detection and side-channel analyses for cyber-security in cyber-physical manufacturing systems," *Procedia Manufacturing*, vol. 1, pp. 77–85, 2015.

[15] M. Yampolskiy *et al.*, "Security challenges of additive manufacturing with metals and alloys," in *Critical Infrastructure Protection IX*, Springer, 2015.

[16] J. C. Jensen *et al.*, "A model-based design methodology for cyber-physical systems," in *Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE, 2011.

[17] R. Kothamasu *et al.*, "System health monitoring and prognostics–a review of current paradigms and practices," in *Handbook of Maintenance Management and Engineering*, pp. 337–362, Springer, 2009.

[18] M. P. Groover, *Automation, production systems, and computer-integrated manufacturing*. Prentice Hall Press, 2007.

[19] M. Al Faruque *et al.*, "Acoustic side-channel attacks on additive manufacturing systems," in *ACM*, 2016.

[20] H. Debar, M. Dacier, and A. Wespi, "Towards a taxonomy of intrusion-detection systems," *Computer Networks*, vol. 31, no. 8, pp. 805–822, 1999.

[21] A. Cardenas *et al.*, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, p. 5, 2009.

[22] B. Evans, *Practical 3D printers: The science and art of 3D printing*. Apress, 2012.

[23] J. Hendershot, "Causes and sources of audible noise in electrical motors," in *Incremental Motion Control Systems and Devices Symposium*, 1993.

[24] S. Yang, *Low-noise electrical motors*, vol. 13. Oxford University Press, USA, 1981.

[25] L. T.-P. Timár-P and P. Tímár, *Noise and vibration of electrical machines*, vol. 34. North Holland, 1989.

[26] J. F. Gieras *et al.*, *Noise of polyphase electric motors*. CRC press, 2005.

[27] "k-quad 5.1 250mm quadcopter frame." thingiverse, http://www.thingiverse.com/thing:397036, 2014.