Quality assurance and situational awareness are important areas of interest for CSIRTs and security teams. Significant efforts have been made on defining metrics measuring critical parameters for these fields of application. However, methodical approaches are missing or lacking precision to enable a reliable usage of such metrics for quality assurance and situational awareness.

In this contribution, we introduce a method that generalizes the application of ARIMA time series analysis on a well-defined set of metrics (ARIMA supplemented metrics) to facilitate and support quality assurance and situational awareness services. This method is based on research on ARIMA models and metrics and builds on CSIRT best practices. We show how data analysts and security practitioners can incorporate this method into existing best practices for CSIRT services pertaining to quality assurance and situational awareness.

The applicability of this method is demonstrated by integrating ARIMA supplemented metrics into exemplary processes for quality assurance and situational awareness to support data analysts and security practitioners in CSIRTs and security teams.

CCS Concepts: • Security and privacy → Malware and its mitigation; Intrusion detection systems; • Mathematics of computing → Time series analysis; Exploratory data analysis; Stochastic processes; • Security and privacy → Denial-of-service attacks; Formal security models;

Additional Key Words and Phrases: ARIMA supplemented metrics, CSIRT services, threat intelligence, situational awareness, quality control, ARIMA, prediction based anomaly detection, time series analysis

ACM Reference format:
Jan Kohlrausch and Eugene A. Brin. 2020. ARIMA Supplemented Security Metrics for Quality Assurance and Situational Awareness. Digit. Threat.: Res. Pract. 1, 1, Article 6 (March 2020), 21 pages.
https://doi.org/10.1145/3376926

1 INTRODUCTION AND MOTIVATION

Collection and sharing security information play a crucial role for CSIRTs and security teams such as Security Operation Centers (SOCs). A comprehensive overview of sources of security data and their applicability for CSIRT services is provided by ENISA in Ciobanu [13]. In addition to that, ENISA highlights the importance of information sharing of actionable threat intelligence information for CSIRTs and security teams in ENISA [12].

Threat intelligence data is shared or offered in different ways and methods. First, technical data is provided by a large number of public and commercial data feeds (e.g., [22]) containing data pertaining to different forms
of malicious activity (e.g., botnets, spam email distribution, known compromised systems). A recent initiative is MITRE’s “ATT&CK” framework, which collects and categorizes a large number of specific techniques that are used by attackers [8]. Moreover, technical threat intelligence systems such as MISP (e.g., [37]) and the “Central Clearing House” resulting from the EU project ACDC (Advanced Cyber Defense Centre, [1]) are actively deployed platforms sharing security incident data, attack techniques, and indicators of compromise (IoC).

Research identifies multiple areas of interest pertaining to sharing and processing security data such as data provided by threat intelligence feeds and network security data. A large number of contributions [16, 22, 25, 28, 32] stress the importance of quality assurance for sharing and using threat intelligence information. Specifically, Li et al. [22] measure and discuss their work the quality parameters to evaluate the quality of data feeds. Important research is done in other works [7, 29, 39, 41] on using data of threat intelligence sources for situational awareness and anomaly detection. According to the Forum of Incident Response and Security Teams (FIRST), ”Situational awareness includes being aware of the current state, and identifying or anticipating potential changes to that state.” Moreover, anomaly detection is characterized in FIRST [27] as “the systematic and often directed searching for anomaly activity inside and outside of network boundaries based upon external and internal information and trends.” These contributions apply time series analysis on the data for measuring the “current state” and to detect changes referring to anomalies. The Microsoft APT Defender team [33] emphasizes the role of anomaly detection for situational awareness.

A comprehensive list of CSIRT services pertaining to information security is provided by the CSIRT services framework published by FIRST [27]. Since FIRST can be considered as the primary organization linking all worldwide CSIRTs together (e.g., [38]), the services framework can be thought of as a standard for CSIRT services. In addition to that, the Handbook for Computer Security Incident Response Teams (CSIRTs) [38] is a best practices document on how to operate a CSIRT and how to define CSIRT services. An important finding is that situational awareness and quality assurance are highly relevant for CSIRTs to facilitate information security services. Situational awareness is explicitly suggested in FIRST [27] as a service area for CSIRTs. The CSIRT handbook proposes to ensure that quality assurance is an essential part of each service definition of CSIRT services.

Metrics as defined by NIST in Jansen [20] are an important instrument in information security. The previously introduced research applied metrics to facilitate quality assurance and situational awareness by quantifying data properties (e.g., measuring quality parameters of a threat intelligence feed). Moreover, metrics and associated measurements are also an important instrument for SOCs and CSIRTs (e.g., [42]). To support applications of metrics on security and threat intelligence data, best practices and guidelines are publicly available (e.g., [14, 21, 23]). As identified by ENISA in Trimintzios [35] and NIST in Jansen [20], uncertainties may affect measurements taken by a metric. To address uncertainties and time dependencies in the measurements, the contributions in other works [29, 39, 41] applied the ARIMA approach of time series analysis. As detailed in Cryer and Chan [10], ARIMA and related approaches are applied in statistics to mathematically model the behavior of time series. In such approaches, time series are dissected into deterministic and pure random components. This allows ARIMA models to facilitate the detection of unexpected measurements by forecasting future data. We refer to Section 3 and Appendix A for a detailed introduction to ARIMA.

1.1 Research Questions

Based on the relevance of quality assurance and situational awareness for CSIRT services and following research achievements on security metrics in such fields of application, we will address the following research questions in this contribution:

• ARIMA proved to be beneficial for specific security metric applications to situational awareness to forecast measurements and to detect anomalies. However, can this method be generalized on a generic set of security and quality metrics?
• Our intuition is that CSIRT services that either contain quality assurance processes or facilitate situational awareness benefit from building on metrics supplemented by ARIMA models. Can we identify relevant services and validate the usefulness of this class of metrics?

Following the design science paradigm [17, 24], our contribution provides, discusses, and evaluates a method for security practitioners, data analysts, and researchers in CSIRTs and SOCs to enable situational awareness and quality assurance. Both situational awareness and quality assurance are supported by supplementing metrics with the ARIMA stochastic approach for data forecasting and anomaly detection. The resulting method is designed to be applicable to a well-defined set of proposed security metrics to provide a foundation for decision-making processes as required for quality assurance and in information security services of CSIRTs.

We apply the design science paradigm to guide the research process. Its design process enables the development and evaluation of IT artifacts through a systematic search process (see Hevner et al. [17] and March and Smith [24]). The artifact we deliver encompasses a method and its instantiation. The resulting IT artifact is evaluated with respect to its utility for CSIRT services in the domains of quality assurance and situational awareness.

2 OVERVIEW OF SECURITY METRICS

There are many different proposals and definitions of what a security metric is and what the fields of application are. According to Black et al. [2], the principal aim of a metric is to “facilitate decision making and improve performance and accountability.” Moreover, Cheng et al. [7] add the technical interpretation that “Security metrics can be naturally interpreted as a standard (or system) used for quantitatively measuring an organization’s security posture.” We herein rely on the organizational aspects that metrics have to support decision making and technical understanding that security metrics quantify the properties of data pertaining to information security. According to FIRST [27], this is an important aspect of data collection facilitating situational awareness: “Overall the information helps provide actionable information to aid in decision making and incident handling” (Section 8.1.3).

Contributions to security metrics fall in different categories that encompass “standardisation efforts such as ISO/IEC 27001,” “Best practices in information security,” and “research.” In the ISO 27001 standard family, metrics are an important instrument measuring the effectiveness of an information security management system. Best practices provide guidance on the selection, definition, and application of security metrics that includes the provision of sets of recommended security metrics. Best practices documents are published, for example, by non-profit organizations such as the Center for Internet Security (CIS) [14] and the SANS Institute [7]. Looking at published metric guides such as CIS [14], the majority of metrics either count observations or average the measured numbers in a specific time interval (e.g., mean time of handling an incident). Thus, our method as proposed in this contribution has to address both types of metrics. Significant achievements in research have been made by applying security metrics in the fields of measuring quality parameters and cyber situational awareness by detecting anomalies. For example, security metrics are the method of choice measuring the quality of threat intelligence feeds in Li et al. [22]. Furthermore, the application of security metrics for cyber situational awareness is intensively discussed in Cheng et al. [7].

Directions of future research on security metrics are proposed by NIST and ENISA in Trimintzios [35, 36]. Both publications recommend conducting research on the measurement as applied by metrics. NIST proposes this in Jansen [20]: “The research goal is to establish formal models with a level of detail sufficient to enable realistic predictions of operational system behavior and portray security measurements accurately.” Furthermore, ENISA discusses in Trimintzios [36] the “uncertainty when measuring security/resilience metrics” and explicitly recommends for resilience metrics the “use of probabilistic measures and models” to consider “more accurate computation of deviations and errors involved” (p. 27). In addition to that, stochastic approaches of time series analysis are recommended as a model used in the analysis of the measured quantities. Although this is
recommended for resilience metrics, we strongly believe that this should also be applied on security metrics in general. Other sources such as Jelen [21] underline that measurements as being taken by metrics are a critical point and that a metric should not rely on an isolated measurement. Furthermore, stochastic approaches have been successfully applied on threat intelligence data to address uncertainties in the data (e.g., [29, 39, 41]). Accordingly, we base our method on research and best practices of metrics.

3 INTRODUCTION TO THE ARIMA STOCHASTIC APPROACH

A well-known approach of time series analysis is provided by ARMA (Auto Regressive Moving Average) modeling time series as linear stochastic processes (e.g., [10]). In particular, ARMA time series analysis builds on an autoregressive (AR) and a moving average (MA) process to address specific mathematical properties of time series. An important property of ARMA processes is that these models are specifically intended to fit stationary time series, which in simplified terms applies to all time series whose mean and noise do not change over time (e.g., see Cryer and Chan [10] for more details). Because ARMA is not explicitly designed to model time series that have, for example, a trend or seasonal component, a generalized approach, namely ARIMA, exists to fit non-stationarity time series. In particular, ARIMA provides means of transforming a non-stationary time series (e.g., time series with a linear trend) into a stationary one to be applicable for ARMA.

In research, ARIMA models have proven to be an effective approach facilitating situational awareness on network security data. Yong et al. [40] introduced an approach to “network security situation awareness” applying ARIMA models for “situation prediction” based on time series of attack rates. Other research contributions, including Rodríguez and de los Mozos [29], Yaacob et al. [39], and Zhou et al. [41], underline the applicability of ARIMA to model properties of network traffic and security data to detect anomalies in such data.

3.1 Applicability of ARIMA on Security and Quality Metrics

The research contributions cited previously have proven the usefulness of ARIMA for specific network security data. Although the term metric is not explicitly mentioned in these contributions, these approaches rely on a quantification of data that aligns to the definition of metrics (e.g., [20]). Thus, ARIMA has been successfully applied to certain security metrics.

Based on the sets of proposed security and quality metrics in CIS [14] and Lowans [23] and the aforementioned research contributions, we are confident that supporting metrics by ARIMA time series analysis can be generalized to a well-defined set of proposed security and quality metrics. This is based on the following observations:

- A large number of common security and quality metrics count observations. These metrics naturally produce results that can be represented as time series allowing them to be modeled by ARIMA. However, specific consideration should be given to the sampling frequency (time interval for quantification of observations). Furthermore, it is shown in Section 4 that other metrics computing a mean (e.g., mean recovery time of incidents) can also be brought into a suitable form.
- The research contributions in Rodríguez and de los Mozos [29], Yaacob et al. [39], Yong et al. [40], and Zhou et al. [41] rely on the characteristics of a time series and do not require, for example, any details about the structure or semantics of the data. Thus, it is plausible to assume that the applicability of ARIMA time series analysis applies not only to the aforementioned specific use cases but can be generalized to assessing measurements taken by a diverse set of metrics.
- Our assumption is confirmed by recommendations in Trimintzios [35, 36] and Jansen [20]. ENISA explicitly recommends in Trimintzios [36] to apply approaches of time series analysis to address uncertainties in the measurements.

\[ \text{It is important to note that ARMA models are stationary if they do not have a “unit root” (e.g., [10]). In most applications of ARIMA, models having unit root are unwanted, and these models are excluded by statistical tests such as the Dickey-Fuller test (e.g., [10]).} \]
Furthermore, we believe that the mathematical properties of ARIMA models align very well with our requirements for the analysis process of security and quality metrics:

- **ARIMA provides means to validate its applicability on a metric for a specific time interval**: The applicability of an ARIMA model to a specific time series can be validated (by the Box-Jenkins method in Box and Jenkins [3]). Thus, it is possible to mathematically prove if a metric can be supplemented by ARIMA (method intrinsic validation of its applicability). In this case, ARIMA provides mathematical sound estimates for future measurements and the confidence of the prediction.

- **ARIMA allows predicting future measurements to detect anomalies**: The mathematical model enables us to forecast future measurements and to estimate the accuracy of the results. This allows detecting unexpected or unusual measurements (anomalies) resulting, for example, in an escalation procedure. In particular, ARIMA provides an estimate for the confidence with which an anomaly is detected. It is important to note that the decision-making process based on detecting such anomalies can take that confidence as an objective quality parameter into account. According to the definition in Jansen [20], facilitating decision making is a principal aim of metrics.

- **ARIMA allows baselining**: ARIMA models provide information about the estimated mean and uncertainties (variance) of a time series that do not change over time (stationarity). This data can be used as a baseline for a metric in a certain time interval. Thus, information gain for a data analyst is that changes in the results of a metric are purely random and do not indicate a change caused by a considerable external cause (e.g., number of IDS alerts are randomly fluctuating and are not influenced by an external security issue).

- **ARIMA allows detection of trends**: The approach enables detection of trends in the data. First, this is an important contribution to understand the properties of a metric. For example, if incident costs are trended over time, it is reasonable to expect a further change in future incident costs. It is important to note that averaging a trended series of measurements does not provide a precise estimation of costs of future incidents.

Accordingly, we believe that the ARIMA supplement improves precision and meaningfulness of measurements as provided by a well-defined set of security and quality metrics. Moreover, these advantages contribute in our intuition to the overall goal of metrics “facilitating decision making and improving performance and accountability.”

### 4 Supplementing Security Metrics with ARIMA

Based on the findings in sections 2 and 3, we introduce a process to supplement well-defined sets of security metrics with ARIMA time series analysis. In the following, we refer to this new class of metrics as "ARIMA supplemented metrics". It is important to note, that the process can to a large extent be automated to support the evaluation of a large number of metrics and associated data sources in parallel. Moreover, we show in this section, how ARIMA supplemented metrics facilitate analysing the results of metrics to assess data properties and detect anomalies.

#### 4.1 Process Supplementing Security Metrics with ARIMA

In the following, we specify a process for supplementing a security or quality metric with a mathematical model resulting from ARIMA. This process embraces the definition of a metric, selection of an initial dataset for model learning and fitting, validation of applicability of the ARIMA approach, choice of a specific ARIMA model, and application and re-calibration of that model. The overall aim is to find an optimal ARIMA model that supplements a security metric on a suitable set of data. In the following, we detail each of these sub-processes. See Figure 1 for reference.
4.1.1 Phase 1: Initial Selection of a Metric and Dataset. The first step is to select a threat intelligence metric and a dataset that is well suited for defining and fitting an ARIMA model.

As mentioned previously, all metrics are applicable that either count observations or compute mean values of measured observations. However, the definition of the metric has to consider the following additional points:

**Measurement:** Since the ARIMA approach applies to time series as underlying data structures, the measurement of a metric must be adjusted to produce results in such form. This applies naturally to all security metrics that count observation in specific time intervals (e.g., the number of IDS alerts per day). Metrics that average numbers of observations or measurements (e.g., "Mean Cost of Incidents") are also applicable if the measurements can be represented as a time series (e.g., incidents costs per week). As detailed in the following, the mean is provided for this class of metrics by the parameters of the fitted ARIMA model.

**Time interval:** A time interval has to be specified for each metric in which observations are collected. For example, the length of the time interval is 24 hours for the metric “IDS alerts per day.” The time interval should be large enough that the time series is not sparse (does not contain frequent "0" measures) but short enough to produce a time series of an appropriate length as defined in the following. For network security data, an interval of 24 hours is often reasonable. Moreover, this interval is used by the most common threat intelligence data feeds, such as the feeds offered by the SANS Internet Storm Center (ISC) [18].
The initial dataset is used in this process to determine a suitable ARIMA model and should comply with the following requirements:

**Representative:** The initial dataset should be representative in terms of a baseline for the applied metric. For example, an initial dataset for producing an ARIMA metric for “network traffic volume per day” should not contain significant changes in that data (e.g., change of trend). Furthermore, the dataset should not contain significant or frequent gaps. For that reason, we propose to manually select an initial dataset or refine the specification of measurement (e.g., time interval) if the process fails to produce a suitable ARIMA model.

**Length:** As a rule of thumb, the length of the time series resulting from the metric should be about 50 elements or longer to prevent overfitting of the model parameters (e.g., see Box and Jenkins [3], Box and Tiao [4], and Shumway and Stoffer [31]). There is no mathematical constraint in applying ARIMA modeling to shorter time series (as far as the ARIMA model parameters can be fitted, e.g., [10]). However, because the accuracy with which the parameters of an ARIMA model are fitted statistically increase with the length of the time series, this may lead to a lack of precision for forecasting values (e.g., [10]).

After the selection of the metric and a dataset, a time series is produced by applying the metric computation schema to the data.

4.1.2 **Phase 2: Refinement of the Dataset.** The goal of this phase is to ensure the applicability of the ARIMA model family. The ARIMA approach imposes specific requirements on the processed time series. In particular, the time series has to be either stationary or stationarity must be achievable by differencing. To ensure the applicability of the ARIMA model family, it is validated by if the time series is either stationary or if it can be transformed into a stationary time series by differencing.

Most applications of ARIMA limit the order of differencing to 2 (see Cryer and Chan [10] for further details). If the current order of differencing exceeds 2, another option to make time series stationary is to further limit the interval of measurements or use a different time interval for the measurements. Sometimes a time series changes its properties or behavior over time. In this case, it is important to limit the time series used to fit the ARIMA model to an interval in which the results are as stable as possible.

4.1.3 **Phase 3: Definition of an ARIMA Model.** The next step is to find an instance of the ARIMA($p$, $d$, $q$) model family that optimally adapts to the data and the previous test for stationarity. $p$ and $q$ is the order (number of dependent variables) of the AR and MA models, respectively. $d$ is the order of differencing (number of recursive differencing operations on the time series). The concrete value for $d$ results from the test of stationarity where $d = 0$ is valid for stationary time series. If a stationary time series can be produced by differencing one or two times, then $d = \{1, 2\}$ dependent on the order of differencing. As mentioned previously, we limit $d$ to $\{0, 1, 2\}$ in our current implementation to prevent model complexity.

In addition to defining a suitable ARIMA model, this process reveals valuable properties for data analysts to interpret metric results:

- **Mean and variance:** The fitted ARIMA model provides the estimated mean of the time series and the estimated variance of the uncertainties of measurements.
- **Trend:** Is there a trend in the results of the metric? A trend indicates that the results of a metric change over time. For example, this provides information if an attack method grows in popularity over time. Moreover, the effectiveness of a security control can be validated by detecting a trend in the results (e.g., falling number of IDS alerts).

---

2 An important presumption of the ARIMA approach is an IID Gaussian noise process. A sparse time series commonly violates the presumption of such a Gaussian noise process (e.g., see Shumway and Stoffer [31]).

3 The AR and MA parts of ARIMA impose stationarity on time series. The integrated (I) component corresponds to the degree of differencing that is required to transform the original time series into a stationary one by differencing.
• **Level of stationarity:** The stationarity is an important indicator for baselining the results of a metric. A stationary time series implies that a time-invariant baseline can be computed. This is especially relevant for all metrics that compute the mean of multiple measurements. For such metrics, a stationary time series is a mathematical sound prove that the mean value is representative.\(^4\)

• **Correlation between metric results:** The values of the AR and MA parameters indicate whether multiple consecutive results of a metric are independent of each other or if one result influences other results.

### 4.1.4 Phase 4: Application of the ARIMA Model

**Estimation of the mean.** As mentioned earlier, a class of metrics measures the mean of multiple measurements. For such metrics, the mean results directly from the fitted parameters of the ARIMA model. Furthermore, ARIMA provides an estimation of the uncertainty of the measurements (e.g., variance of incident costs) and a forecast of future data (e.g., expected costs of new incidents). It is important to note that forecasted values by an ARIMA supplemented metric take properties such as trends and correlations of measurement into account. For that reason, ARIMA supplemented metrics produce in our intuition more meaningful and precise results compared to metrics that solely compute an arithmetic average of measurements (e.g., \([\text{10}]\)). For example, if the costs of security incidents increase over time, the average cost of incidents is not suitably predicting future costs precisely. Furthermore, we argue that decision-making processes based on metric results clearly benefit from these advantages.

**Anomaly detection.** In this phase, the fitted ARIMA model is applied for prediction-based detection of anomalies and outliers. Anomaly detection is characterized in FIRST \([\text{27}]\) as "the systematic and often directed searching for anomaly activity inside and outside of network boundaries based upon external and internal information and trends." Here, we understand any single unexpected measurement or a fundamental change of the properties (e.g., change of the mean or trend) of a metric as an anomaly.

The approach for anomaly detection in ARIMA supplemented metrics builds upon the detection of outliers in time series (e.g., \([\text{10}]\)). As defined in Cryer and Chan \([\text{10}]\), outliers refer to atypical observations caused by measurement error or abrupt changes in the underlying process. We take the following semantic distinction between outliers and anomalies:

**Anomaly:** A specific event (e.g., emerging threat or novel attack strategy) that causes a significant change in the measurement of a metric.

**Outlier:** Atypical and unexpected observation in the time series caused by measurement error or abrupt change in the underlying process. Here, we refer to an outlier as a mathematically well defined instantiation of an anomaly.

Since outliers are well defined for ARIMA time series analysis\(^5\) and methods are available to detect outliers \([\text{10}]\), we now leverage the method for detecting outliers to identify anomalies in the results of a metric.

Detection of outliers relies on a prediction of expected future measurements and estimated confidence intervals (e.g., \([\text{31}]\)). In particular, we label a result as an outlier if the measured result is outside the predicted 95% confidence interval. It is important to note that outliers may be caused by an abnormal large error in the measurement. Thus, there is no corresponding anomaly that is caused by external activity ("false positive"). To validate and assess anomalies, an additional process is required. Since we expect an anomaly to be caused by an external event or incident, it is crucial to check whether other sources of threat intelligence information also provide evidence for this anomaly in the same time interval (time-based correlation):

\(^4\)For a Gaussian distributed random variable or process, the mean is the best estimate for the expected value.

\(^5\)We refer to Cryer and Chan \([\text{10}]\) for a mathematical definition of outliers as part of the ARIMA approach.

Digital Threats: Research and Practice, Vol. 1, No. 1, Article 6. Publication date: March 2020.
Other threat intelligence platforms such as MISP provide information about novel attack methods or threats. For example, a rise in IDS alerts may coexist with reports of a new threat or attack method.

Multiple different data sources may exhibit a similar characteristic. For example, the number of IDS alerts may simultaneously rise together with the volume of network traffic or blocked network connections.

Log files of applications may reveal unexpected entries that fall in the same time interval of observation (time-based correlation). An example of such a correlation is provided in work of the Microsoft Security Response Center [6].

The reliable validation of suspected anomalies by correlating information of different threat intelligence sources is difficult to automate. For that reason, in our current, process validation and interpretation of suspected anomalies are performed by data analysts.

4.1.5 Recalibration of the Model. Periodic recalibration is required because of different causes:

Anomalies: The resulting time series of measurements may exhibit significant anomalies. Herein, an anomaly is significant if multiple consecutive results of a metric are affected by the anomaly. An example of such an anomaly is a level change where the mean of the time series changes.

Statistical properties: Statistical properties are trends, correlations, and the noise process. Since all of these properties may change over time, a periodic recalibration of the model is required.

4.2 Prototypical Implementation

A prototypic implementation of the process as shown in Figure 1 exists that is based on the programming language Python and the statistics modules “statsmodels” and “pandas.” These Python modules provide methods fitting ARIMA models, receiving the fitted parameters of the model, and statistical tests such as Dickey-Fuller to validate the applicability of a model on the underlying time series. The process is fully automated from the computation of the time series to the detection of outliers. The selection of metric and initial dataset is done manually by data analysts based on the specific goal pertaining to data quality criteria. As detailed earlier, the distinction between outliers and anomalies and the final confirmation that an outlier corresponds to an anomaly requires inspection by data analysts.

In our implementation, we apply the augmented Dickey-Fuller unit root test of Python “statsmodels” to ensure that the initial or refined time series for fitting an ARIMA model is stationary. As mentioned previously, if the test fails, it is repeated on the differences of all contiguous pairs of values (differencing) to produce a stationary time series.

We apply the Python modules “statsmodels” and “pandas” to fit the model parameters to a specific ARIMA model, to forecast values, and to compute the confidence intervals. Currently, we limit all parameters \((p, i, q)\) to \(0, 1, 2\) to prevent overfitting of the model. Furthermore, “statsmodels” provides the values of the fitted model (“params” method) and additional information pertaining to the accuracy and confidence of the model parameter estimation (Python class “ARIMAResults”).

We decide to use the common 95% confidence interval for anomaly detection. A candidate for an anomaly is detected if the measured value is outside of this interval. Although more strict confidence intervals would lead to a lesser rate of false positives, the likelihood to miss an anomaly would be higher. Thus, the 95% confidence interval can be considered as a reasonable trade-off between detection rate and false positives.

4.3 Validation of the Applicability of ARIMA Supplemented Metrics

Validation of the usefulness of ARIMA supplemented metrics requires multiple co-dependent steps. Although there are the previously listed mathematical advantages of the ARIMA approach, it is still unclear if the approach proves to be applicable and advantageous for the fields of application and targeted CSIRT and security community. As mentioned earlier, targeted fields of application are quality assurance and situational awareness.
facilitating and supporting CSIRT services. Since best practices pertaining to these fields of application are provided by FIRST [27, pp. 41ff] and the CSIRT handbook in West-Brown et al. [38, pp. 42ff], our validation is based on these contributions.

For the validation of the applicability of the method, we propose an evaluation of two hypotheses:

**H1 (Relevant CSIRT data sources fit ARIMA models):** ARIMA models fit current threat intelligence data. To show this, we identify feed data that has been already used in other contributions and that we believe to be representative. Criteria for a representative feed is the reputation of the source and the size and importance of that feed.

**H2 (ARIMA supplemented metrics are applicable for CSIRT use cases):** Based on the selection of sources, we will indicate that this new class of ARIMA supplemented metrics produces the predicted precise and meaningful results for specific applications supporting exemplary CSIRT services. We selected the use cases because they point out from our experiences to be important fields of applications for CSIRTs and security teams.

In the following, hypothesis H1 is proved by validating that ARIMA is applicable to common and relevant sources of threat intelligence data. Hypothesis H2 is validated by integrating ARIMA supplemented metrics into CSIRT services based on FIRST [27] and West-Brown et al. [38]. Since the introduced method is generally applicable to a large set of security metrics and is explicitly designed to support diverse fields of application with respect to data and technical systems, an exhaustive experimental evaluation is outside the scope of this work. Instead, we highlight its achievements for quality control and situational awareness based on previous research and its mathematical properties and focus on exemplary use cases for empirical validation. However, future plans exist to reach out and incentivize the international and European CSIRT communities (e.g., FIRST and TF-CSIRT) to further investigate fields of applications and validate the usefulness of ARIMA supplemented metrics for CSIRT services.

5 APPLICATION TO DATA QUALITY ASSURANCE

Data quality control and assurance is an integral part of information security and addressed by research, best practices, and standardization initiatives. In Cykana et al. [11], different dimensions of data quality criteria are introduced, which comprise accuracy, uniqueness, timeliness, validity, completeness, and consistency. Standards such as ISO 9000 and ISO 27001 families [5, 34] are dedicated to quality management systems in information security. ISO 9000 states that quality assurance can be defined as the “part of quality management focused on providing confidence that quality requirements will be fulfilled” (e.g., [15]). This definition applies to both products and services. According to ISO 9000, quality control focuses on the inspection aspect of quality management and not on the complete process. In the following, we use the term *quality assurance* because this term is more comprehensive in respect to quality management.

A significant number of research contributions, which include Grispos et al. [16], Meier et al. [25], and Sillaber et al. [32], underline the importance of quality assurance and control for threat intelligence applications. In addition, Li et al. [22] and Pinto and Sieira [28] successfully apply security metrics to measure quality criteria of common feeds of threat intelligence data. Quality criteria relate in both contributions to either a single feed or a pair of feeds (e.g., intersection of IP addresses). Li et al. [22] additionally provided an approach to measure the accuracy of threat intelligence data feeds. An important finding is that the relevance of certain quality criteria depends on the specific use case. As an example, the authors argue that the latency of IoC is important for attack detection, but it is not critical for post-mortem forensic investigations.

In addition to research contributions, quality assurance is an essential part of the CSIRT services “Incident Handling” and “Situational Awareness” as defined in the CSIRT handbook [38] and the CSIRT services framework [27]. Pertaining to situational awareness, the CSIRT services framework emphasizes that “Sources of cybersecurity data need to be validated for accuracy often due to a high number of false positives” (Chapter 8.1.4,
“Function: Data processing and preparation”). The CSIRT handbook dedicates a chapter to quality assurance and proposes to define and use a quality assurance system (QAS) as part of CSIRT service definitions. The QAS consists of the parts “definition of a quality system,” “checks,” and “balances.” The first part of the QAS aims at defining “quality parameters” for being able to satisfy quality assurance levels defined by the CSIRT mission statement. The second part “check” demands to measure the previously defined quality parameters: “Having defined quality parameters, one also needs to define how to check these parameters and how to measure them. This is by no means a trivial task and dictates some serious a priori measures” (Chapter 2.6.2, p. 45). It is important to note that this part of the QAS aligns with our proposed method, using security metrics for assessing data quality properties. For that reason, we argue that ARIMA supplemented metrics may be beneficial in implementing that part of the QAS. The final part aims at defining procedures assuring quality including crises management: “Procedures must be in place to enforce quality when it is at risk” ([38], p. 47). Since ARIMA supplemented metrics allow detection of unexpected drops in quality parameters, we believe that our method can be applied supporting that task.

5.1 Data Quality Assurance Process Based on ARIMA Supplemented Metrics

Although research provides evidence that ensuring data quality is crucial for threat intelligence data and that security metrics provide an efficient instrument measuring data quality, we identified demands for further research:

- Important work measuring the quality of threat intelligence feeds is contributed in Pinto and Sieira [28] and [22]. However, both contributions do not take the context (e.g., security control) into account in which the data is supposed to be applied. We believe that complementary work should additionally measure the effectiveness of using such data feeds resulting in security improvements. For instance, a specific security improvement resulting from deploying a security control might be verified by time series analysis detecting a drop in the IDS alerts or AV alerts.

- The measurement of data quality is done using either isolated measurements or by computing the mean of multiple measurements (e.g., [22]). It is reasonable to assume that conducting a time series analysis based on data feed measurements may produce more precise results.

- As detailed in the CSIRT handbook and services framework, quality assurance should be part of the CSIRT service definition (e.g., for Incident Response). As argued previously, we propose to further investigate how ARIMA supplemented metrics can be integrated into the quality assurance definition and process implementation of CSIRT services. Following the first point, we believe that our method might contribute measures for the effectiveness of security controls by time series analysis. Furthermore, we propose that stochastic approaches such as ARIMA can facilitate defining and identifying crises due to dropped data quality parameters by detecting anomalies in the measured quality parameters.

Accordingly, we propose to apply ARIMA supplemented metrics to achieve these goals. A data quality assurance process as defined in this contribution pertains to the activity of assuring and controlling data quality as defined by a QAS as introduced in West-Brown et al. [38]. We strongly believe that such a QAS can benefit from integrating ARIMA supplemented metrics in the measurement part (“check”) and propose to investigate the advantages pertaining to threat intelligence applications. Since threat intelligence applications such as MISP gain importance for CSIRT services (e.g., sharing of IoC and security events, see ENISA [12]), we argue that QASs also apply to such applications. The overarching aim of the application of ARIMA supplemented metrics is to improve detection of time-dependent changes in the data quality and spot anomalies in the measured data. Following the approach in the CSIRT handbook in West-Brown et al. [38], we propose an adapted process defining a QAS that applies to CSIRT services:
Selection of quality parameters: As specified in West-Brown et al. [38], the first step is to define quality parameters that allow one to validate if the overall quality goals of the services have been achieved. These parameters are supposed to be measured during the quality assurance process. As detailed in the CSIRT handbook, the choice of these quality parameters depends on specific requirements that have to be defined for each service. Furthermore, although specific quality parameters for threat intelligence feeds are provided by Li et al. [22], the authors point out that the importance and criticality of parameters such as “timeliness” vary with the intended use case (attack detection vs. post-mortem analysis). For that reason, it is not practicable to define a universal set of parameters that apply to all use cases. Instead, we propose to follow the generic process of QAS definition as introduced in West-Brown et al. [38].

Definition of suitable quality metrics: As detailed in Section 4, the next step of our method is to define suitable ARIMA supplemented metrics that allow one to measure the quality parameters as selected in the first step. Based on the metric guidance in CIS [14] and Lowans [23] and current research in Li et al. [22], we identified the following ARIMA supplemented quality metrics to guide corresponding use cases:

- **Data volume**: Anomalies in the volume of threat intelligence data (e.g., incidents, IDS and AV alerts). This reveals short-term changes in the data that are, for example, caused by technical issues pertaining to the data handling.
- **Effectiveness**: Change points in the volume of threat intelligence data (e.g., incidents, IDS and AV alerts). This reveals long-term changes caused by attack trends (rise) and security controls (drop).
- **Accuracy**: Trend and mean value of the number of false positives in threat intelligence data.
- **Timeliness**: Trend and mean value of the timeliness of threat intelligence data (e.g., reported incidents).

Definition of procedures to ensure quality: As detailed in the CSIRT handbook, procedures are required to react to a situation where quality is at risk. In the quality assurance process, these situations correspond to the occurrence of anomalies in the measured quality parameters. For each type of anomaly, a specific procedure has to be defined to react to the specific situation in which the anomaly occurs.

5.2 Experimental Results

In the following, we illustrate the data quality assurance process by using a threat intelligence platform operated by an academic CSIRT as an example. The specific aim of the process is to react to rapid changes in the number of ingested security events that may indicate technical problems pertaining to the system itself or a constituent site submitting data. This corresponds to validating hypothesis H2 (Section 4). We omit a validation of hypothesis H1 based on the data of that platform because the data is private. In addition to that, we believe that other sources of threat intelligence information are better suited to transfer the validation process to common data sources and refer to Section 6.2 where this hypothesis is validated in a verifiable way on a public data feed from the SANS ISC.

The threat intelligence platform collects security events such as attacks against honeypots, IDS and AV alerts, spam emails, and events from threat intelligence feeds from diverse sensors as deployed in networks of national Internet service providers (ISPs) and CSIRTs. The platform identifies the source of the attack (e.g., source IP address) and informs the contact that is administratively responsible for the site (e.g., university):

Selection of quality parameters: Our motivation applying the process is to detect technical problems resulting in a rapid unexpected change in the volume of ingested security events. The corresponding quality parameter that aligns with these requirements is the number of ingested security events per day.

Definition of suitable quality metrics: As defined in the first step, the ARIMA supplemented metric is based on the number of ingested security events per day. We decided to start with unlimited test data (time series of event numbers per day) that is applied to generate an ARIMA model. Refinement of the dataset is required if the model cannot be fitted to the unlimited data. Since the numbers of events per day is usually very high, significant drops in the numbers likely reveal technical problems.
Definition of procedures to ensure quality: The time series of security events are displayed by a web application. Detected anomalies are highlighted and marked inside the time series. Currently, further analysis of anomalies is done by CSIRT members monitoring the results of the ARIMA supplemented metric. It is reasonable to assume that technical problems with ingesting data will lead to a massive drop in the number of security events for at least 1 day. For that reason, a short-term anomaly triggers a process to check for problems pertaining to data ingestion. A change point in the data can have multiple different causes. In this situation, it is important to find out whether this change is caused by a global trend in attacks or if it corresponds to changes in the submission of security events by one or more constituent sites. If the change is not caused by a constituent site, further investigation performs correlation with other threat intelligence information (e.g., MISP events) to pinpoint a root cause of the anomaly.

A screenshot of this threat intelligence platform is shown in Figure 2. It can be seen that the anomaly detection is implemented as part of the quality assurance process on the threat intelligence platform. Each suspected anomaly is labeled by a small red dot on the horizontal red line in the graph. Three anomalies marked by a red vertical line have been further selected for a more detailed presentation. It can be seen that the first anomalies are caused by technical problems (drop of event numbers) due to data ingestion problems. It is interesting to note that each of these outages result in a pair of detected anomalies. The last annotated anomaly is a change point of the number of ingested events resulting from a major change at a site delivering data. Overall, it points out that our method reliably detects all significant anomalies where a rapid change in the number of security events occur. This shows the validity of our hypothesis H2.

6 APPLICATION TO SITUATIONAL AWARENESS SERVICES

MITRE defines situational awareness as “essentially, understanding your environment and accurately predicting and responding to potential problems that might occur” [9]. A similar definition tailored to CSIRT services is provided in the FIRST CSIRT services framework [27]: “Situational Awareness comprises the ability to identify, process, comprehend, and communicate the critical elements of what is happening in and around the CSIRT’s
area of responsibility that may affect the operation or mission of its constituency. Situational awareness includes being aware of the current state, and identifying or anticipating potential changes to that state.

The relevance of security metrics for situational awareness is shown in Cheng et al. [7]. Moreover, the ARIMA approach has been successfully applied in several research contributions (e.g., [29, 39, 41]). In addition to that, the importance of anomaly detection for situational awareness is underlined by the Microsoft APT Defender team, where a new malware artifact has been uncovered by detecting abnormal usage of a Windows program (MSHTA.exe) [33]. Time series analysis revealed an emerging campaign spreading this malware.

Situational awareness has also become important for CSIRTs. A comprehensive description of the relevance of situational awareness for CSIRTs and how situational awareness can be offered as a service is provided in the FIRST CSIRT Services Framework [27]. The framework suggests three different service areas: "Data acquisition," "Analysis and synthesize," and "Communication," where the second part demands "analysis techniques to determine what is occurring that may impact the constituency assets and security posture" and "Analysis may reveal when events do not match typical expected behavior." This perfectly aligns to the methodology of ARIMA supplemented metrics to assess data properties and detect unexpected behavior.

6.1 Integration of ARIMA Supplemented Metrics into Situational Awareness Services

Based on the service definition of situational awareness in the CSIRT services framework (chapter 8.2, "Service: Analyze and interpret"), we propose to integrate ARIMA supplemented metrics as a method detecting unexpected behavior in threat intelligence data. A sound integration of ARIMA supplemented metrics into the services framework in FIRST [27] requires consideration of multiple factors:

- **Intended achievement of the service:** The goals of the service must be well defined as addressed by the services framework in FIRST [27]. Based on that, it is important that these goals align with the application of ARIMA supplemented metrics. For example, a service monitoring IDS alerts in a constituent network would look for an unexpectedly high number of alerts, indicating a severe incident on the network.

- **Metric selection:** As detailed in Section 4, suitable metrics encompass all metrics that either count observations or compute the mean of observations in a specific time window. As mentioned earlier, the choice of a metric has to suit the intended achievement of that service. First, the gathered data must be suitable to achieve the aim of that service. Moreover, as detailed in Section 4, the granularity of the measurements (e.g., number of events per 24 hours) and the initial dataset to fit the model must be carefully chosen.

- **Data selection:** Taking current research into account [22, 29, 39, 41], the particular types of relevant data comprise threat intelligence data feeds and network security data such as traffic volume, IDS alerts, firewall logs, and events detected on honeypots and darknets. However, if the volume of collected data is sufficient and if its properties suffice to the mathematical requirements as detailed in Section 4, all other threat intelligence sources are applicable. This applies, for example, to the number of attacks or observations that have been classified with the MITRE ATT&CK framework.

- **Metric application:** As detailed in Section 4, further processes are required to exclude false positives (outliers that are produced by an unexpected high measurement error) and assess the root cause of an anomaly. As detailed previously, we propose to set up a process to correlate reported outliers with other sources of threat intelligence data, which includes external data feeds, security data from private sources such as network monitoring or IDS alert, and threat information from MISP. As detailed in the services framework (8.3, "Function: Communication"), additional processes are required for reporting and assessing detected anomalies (e.g., a new malware artifact spreading on a network).

6.2 Experimental Results

We demonstrate the application of ARIMA supplemented metrics for situational awareness (hypothesis H2) by applying our method to a DShield feed of threat intelligence data as provided by the SANS Technology
Furthermore, we show that the data perfectly suits the mathematical requirements of the ARIMA approach (H1). The ISC operated by the SANS Institute provides publicly available threat intelligence data and is well known in the security community. Threat intelligence data is collected by a large network of sensors (DShield). Furthermore, SANS ISC data has proven its value by being used in relevant research (e.g., [22]). Thus, this data feed is actively used in the research, CSIRT, and security community. In addition, the data is publicly available to validate our findings.

The intention is to support a service for situational awareness that monitors a SANS ISC data feed for unexpected malicious behavior that may, for example, indicate an emerging Internet worm.

### 6.2.1 Intended Achievement of the Service

According to work of the Microsoft Security Response Center [6], a vulnerability (CVE-2019-0708) in the Remote Desktop Protocol (RDP) is supposed to be abused to spread a new Internet worm on the Internet. An important aspect of situational awareness is to monitor suitable sources of threat intelligence data for malicious behavior pertaining to attacks against the vulnerability. Since DShield monitors attacks using a large and distributed network of IDS sensors and honeypots, a new worm spreading on the global Internet by abusing the vulnerability would result in a rapid increase of attacks on the sensor network (e.g., [43]). A specific motivation of CSIRTs is to detect new threats (e.g., Internet worms) as early as possible to issue security warnings and to contain the threat as far as this is technically and organizationally feasible.

The specific aim is to apply ARIMA supplemented metrics to obtain situational awareness on the DShield data feed to detect unexpected behavior indicating such a new Internet worm or another emerging threat.

### 6.2.2 Metric and Data Selection

The characteristic of the appearance of a new RDP worm is an abnormally high number of sources scanning for that vulnerability on port tcp/3389. For that reason, this service has to measure the number of sources scanning for that vulnerability.

The service adopts the following metric from DShield: “number of sources per day connecting to port tcp/3389 on a DShield sensor” [19]. A granularity of 24 hours (number of sources per day) is well suited. On the one hand, this time slice is short enough for a timely detection of a worm. On the other hand, the number of sources per day is large enough to produce a time series that is not sparse and does not contain any obvious gaps.

For fitting an ARIMA model, a time interval of 2 months is used (from April 2019 to June 2019). The time series contains about 60 measurements, which suits the recommended length of a time series for fitting an ARIMA model (e.g., [3, 30]).

### 6.2.3 Application of the ARIMA Supplemented Metric

The number of IP addresses targeting port tcp/3389 (Microsoft RDP) on the ISC sensors is shown in Figure 3. The ARIMA supplementing process results in an ARIMA(2,0,1) model. Since the augmented Dickey-Fuller test (e.g., [26]) on the model parameters confirms that the time series is stationary, the ARIMA supplemented metric Number of connections targeting tcp port 3389 is without refinement applicable on the ISC dataset. Thus, the process facilitating anomaly detection can be applied on the original data feed (Section 4). It is important to note that this satisfies our hypothesis H1.

### 6.2.4 Prediction of Value and Confidence Interval

In Figure 3, the forecasted values are marked by a red dotted line, whereas as the 95% confidence intervals are displayed as green dotted lines. Prediction is dynamically done based on the previously determined ARIMA model parameters and the preceding values of the time series.

The preliminary findings of that service validating H2 are as follows:

**F1:** The data as shown in Figure 3 suits the applied ARIMA supplemented metric without modification (is stationary). This proves that there is no trend or other deterministic change in the results of the metric.

**F2:** It can be seen in Figure 3 that drops in the number of sources produce outliers.
Fig. 3: Number of sources attacking port tcp/3389 on DShield honeypots. Data is gathered from the DShield database (https://dshield.org/port.html?port=3389). The red dotted lines are forecasted values. The confidence intervals are marked by green dotted lines.

**F3:** The forecasted values and estimated confidence intervals in Figure 3 align with our intuition of abnormal measurements. Thus, ARIMA time series analysis is well suited to forecast future measurements and provide plausible results detecting anomalies.

Based on the findings, we conclude the following:

- Finding F1 indicates that the malicious activity pertaining to Microsoft RDP attacks remained constant over the monitored time interval.
- There are drops in Figure 3 that are labeled as outliers (F2). However, our primary indicator for an anomaly caused by a worm or significant threat is that new measurements of the number of sources exceed the predicted confidence interval. For that reason, at the time of this writing there is no evidence for such malicious activity.
- To the best of our knowledge, the vulnerability in the RDP service (CVE-2019-0708) has not been abused by an Internet worm, so far. However, since ARIMA is a well-researched method for anomaly detection (e.g., [39]) and we have confirmed the suitability of the data for ARIMA (F1 and F3), we are confident that our method will detect such a worm if it starts spreading on the Internet.

7 CONCLUSION AND FUTURE WORK

7.1 Conclusion

In this contribution, we introduced, analyzed, and discussed a method supplementing well-defined sets of security metrics with ARIMA time series analysis. We identified situational awareness [7, 29, 39, 41] and quality assurance [22, 25] as important fields of application supporting security practitioners and data analysts at CSIRTs and security teams, as well as researchers concerned with threat intelligence applications. Furthermore, time
series analysis based on ARIMA models has been successfully applied on threat intelligence data and for situational awareness [29, 39, 41]. Based on the best practices for CSIRTs services in First [27] and West-Brown et al. [38], we demonstrated how this method can be integrated into QASs and situational awareness services of CSIRTs and associated security teams (e.g., SOCs).

The method as introduced in this contribution generalizes the application of ARIMA time series analysis on a well-defined set of security metrics (e.g., [14, 23]), resulting in a class of ARIMA supplemented metrics. We presented a process of how to supplement a certain metric with an ARIMA model. In particular, all metrics are suitable that either count events or compute the average of multiple measurements.

Our main motivation for supplementing metrics by ARIMA models are multifold. ARIMA models allow forecasting of future results of metrics, which allows to detect outliers and anomalies in the data (prediction-based anomaly detection). Moreover, ARIMA allows estimation of uncertainties in the measured metric results. Thus, ARIMA provides a measure of confidence for the measurements that in our intuition is beneficial for the goals of metrics as defined by NIST in Black et al. [2]: “facilitate decision making and improve performance and accountability.” Thus, a decision can be taken based on an estimate of confidence for a specific measurement.

Validation is based on two hypotheses. We demonstrated that ARIMA models are applicable on common threat intelligence data feeds such as provided by the SANS ISC (DShield threat intelligence feed), corresponding to our first hypothesis (H1). To provide evidence for the applicability to situational awareness and quality assurance, we demonstrated how to apply ARIMA supplemented metrics on exemplary applications (hypothesis H2).

7.2 Future Work

In this contribution, we applied prediction-based detection of anomalies based on ARIMA. Thus, a measurement taken by a metric is labeled as an anomaly if the measurement is outside of the estimated confidence interval. Prediction-based detection is easy to implement because it does not require any a priori information about anomalies (often anomalies do not occur frequently, and it is hard to measure, e.g., their statistical properties). However, the downside is often a higher rate of false positives or false negatives compared to approaches that rely on a priori information. To further improve anomaly detection, future work could use a more complex approach for anomaly detection based, for example, on stochastic estimates of the characteristics of anomalies, a priori data derived from long-term observations, or a mathematical model for anomalies (e.g., Zou et al. [43] define a mathematical model for an Internet worm).

Other research efforts could be spent on automating correlation of outliers as detected by ARIMA supplemented metrics with other sources of threat intelligence. It is important to note that well-defined approaches exist to determine if time series are correlated (e.g., [10]). This could, for example, be used to determine whether the results of different ARIMA-based metrics are correlated.

Since the original ARIMA model does not consider seasonal changes or external distortions in time series, extensions to this model have been proposed (SARIMA and ARIMAX). Thus, these extended models may provide an improved precision of forecasting measurements. However, applying such models either requires external input (ARIMAX) or a much larger time series (SARIMA) to address seasonal changes (e.g., [10]).

Future work should investigate the general applicability of ARIMA supplemented metrics for CSIRT services and threat intelligence platforms. Since ENISA emphasizes the importance of threat intelligence platforms such as MISP for CSIRTs and SOCs, future work could be done on applying this class of metrics on such platforms operated in the CSIRT community (e.g., MISP instance deployed by FIRST).

A APPENDIX

A.1 Mathematical Details of the ARIMA Model

An ARMA model is composed of an AR and an MA model. Common to both parts is that they mathematically model linear dependencies in time series data. Since both models differ in their properties, we discuss them in two separated sections, starting with the AR models.
AR models. An AR model of order $p$ is defined as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t,$$  \hspace{1cm} (1)

where $\epsilon_t$ is a Gaussian noise process $WN(0, \sigma^2)$ and $\phi_1, \ldots, \phi_p$ are constants in $\mathbb{R}$. The simplest model of the AR family is the AR(1) model (2), considering a direct linear dependency between the lags $t$ and $t+1$:

$$X_t = \phi_1 X_{t-1} + \epsilon_t.$$ \hspace{1cm} (2)

Thus, if we have a significant distortion at $t-1$, this directly affects the result at $t$. The correlation of variables in AR results in a characteristic pattern in the autocorrelation function (ACF). The ACF computes the correlation of different lags in a time series and is defined by

$$R(\tau) = \frac{\mathbb{E}[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2},$$ \hspace{1cm} (3)

where $\mu$ and $\sigma$ are the mean value and the standard deviation of the time series and $\tau$ is the time lag. It is important to note that we assume a stationary behavior, resulting in a constant mean $\mu$.

Direct correlations between variables in the AR model propagate exponentially smoothed through all parts of the ACF. This is prevented by the partial autocorrelation function (PACF) where indirect correlations, such as between variables $X_t$ and $X_{t-2}$ are compensated. For example, the direct correlation between $X_t$ and $X_{t-1}$ in the model AR(1) would result in a PACF that exhibits only a correlation in the first lag ($\tau = 1$), whereas this correlation would propagate damped through all other lags in the ACF ($\tau \geq 2$).

MA models. The MA model of order $p$ is defined as

$$X_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_p \epsilon_{t-p},$$ \hspace{1cm} (4)

where $\epsilon_{t-1}, \ldots, \epsilon_{t-p}$ are Gaussian noise processes $WN(0, \sigma^2)$ and $\phi_1, \ldots, \phi_p$ are constants in $\mathbb{R}$. The simplest model of the MA family is the MA(1) model:

$$X_t = \epsilon_t + \phi_1 \epsilon_{t-1}.$$ \hspace{1cm} (5)

In analogy to the AR models, MA models exhibit a characteristic pattern in the ACF and PACF. For instance, the ACF shows a correlation between $X_t$ and $X_{t-1}$, whereas the other lags are unrelated. The PACF of the MA(1) model exhibits a characteristic exponential decay. It is important to note that the characteristic patterns of AR and MA models are converse.

ARIMA models. The ARMA(n,m) model family is a combination of both the AR and MA models:

$$X_t = \epsilon_t + \sum_{i=1}^{n} \phi_i X_{t-i} + \sum_{j=1}^{m} \theta_j \epsilon_{t-j}.$$ \hspace{1cm} (6)

It is important to note that ARMA models imply a stationary time series. As such, the time series must have a constant mean $\mathbb{E}(X_t)$ and the covariance function $\text{Cov}(X_{t_2}, X_{t_1})$ depends only on the time difference $\tau = t_2 - t_1$ to satisfy the conditions of weak stationarity.

Generalization of the ARMA model allows to overcome the limitations of the weak stationarity requirement. The ARIMA model is under some circumstances able to convert a non-stationary time series into a stationary one by differencing $n$ times. For example, a linear increase can be eliminated by differencing one time. Other extensions such as ARIMA-GARCH and ARIMAX address varying noise processes and exogeneous variables (external distortions). Other extensions exist to model seasonal behavior in time series.
A.2 Model Identification and Parameter Fitting

Unfortunately, there is no simple and straightforward approach to select a model and its parameters, which makes the choice neither easy nor obvious. Instead, multiple different approaches including the Akaike Information Criteria (AIC) and the Box-Jenkins method have been proposed. AIC is a measure for the relative quality of the model that maximizes a trade-off between the likelihood of the model fitting and the complexity of the model (refer to Shumway and Stoffer [31] for more information). Although the accuracy of a model may benefit from more parameters, it may lead to an overfitting of the model. That is why AIC prefers a less complex model.

The Box-Jenkins method takes the ACF and PACF of the time series into account. As shown previously, each MA and AR model has a characteristic pattern regarding the ACF and PACF. For example, the number of related lags in the PACF indicates the order \( p \) of an AR model. A measure for the applicability of a selected ARMA model and its parameters is given by the ACF of the residuals, which are the differences between the predicted values of the chosen ARMA model and the current time series. It is important to note that ARMA models split up the time series into a deterministic stationary process and a random noise process. An appropriate model is able to correctly predict timely correlations in the time series, which eliminates all timely correlations in the residuals. Thus, an appropriate model leaves in theory all values of the ACF and PACF residuals uncorrelated.

To compute the parameters for the ARMA models, the Python libraries Pandas, statsmodels, and matplotlib have been used. As there is no simple and straightforward approach to select the appropriate model, a test and trial approach can be used that is based on the AIC and Box-Jenkins method. For our work, we used the Python library statsmodels to fit the parameters of the selected ARMA model, to compute the residuals, and to predict future values. As mentioned earlier, we proved the applicability of selected ARMA models and its parameters by the ACF and PACF of the residuals. If more than one model was applicable, we sorted out the optimal model by using the AIC. It turned out that the ARMA(2,1) model

\[
X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta w_{t-1} + w_t
\]

and the ARMA(2,0) model

\[
X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + w_t
\]

provide a good compromise between complexity and accuracy for the our threat intelligence data. Although more complex models may provide a slightly better theoretical error deviation, the results could suffer from overfitting.

ACKNOWLEDGMENTS

We are immensely grateful for the insightful comments and guidance offered by the anonymous peer reviewers at ACM DTRAP and our colleague Christian Keiland explicitly thank the Forum of Incident Response and Security Teams (FIRST) for their support in publishing this work. We would like to thank the SANS Technology Institute Internet Storm Center for making their data available to us.

REFERENCES

[1] ACDC Project Consortium. 2015. Advanced Cyber Defence Centre (ACDC). Retrieved February 1, 2020 from https://www.acdc-project.eu/.
[2] Paul Black, Karen Scarfone, and Murugiah Souppaya. 2008. Cyber Security Metrics and Measures. Retrieved February 1, 2020 from DOI: https://doi.org/10.1002/9780470087923.hhs440
[3] George E. P. Box and Gwilym M. Jenkins. 1976. Time Series Analysis: Forecasting and Control. Holden-Day.
[4] G. E. P. Box and G. C. Tiao. 1975. Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association 70 (1975), 70–79. DOI: https://doi.org/10.1080/01621459.1975.10480264
[5] Alan Calder. 2009. Information Security Based on ISO 27001/ISO 27002: A Management Guide—Best Practice. Van Haren Publishing.
[6] Microsoft Security Response Center. 2019. Prevent a Worm by Updating Remote Desktop Services (CVE-2019-0708). Retrieved September 30, 2019 from https://msrc-blog.microsoft.com/2019/05/14/prevent-a-worm-by-updating-remote-desktop-services-cve-2019-0708/
6:20  •  J. Kohlrausch and E. A. Brin

[7] Yi Cheng, Julia Deng, Jason Li, Scott A. DeLoach, Anoop Singhal, and Xinning Ou. 2014. Metrics of Security. Springer International, 263–295. DOI: https://doi.org/10.1007/978-3-319-11391-3_13
[8] MITRE Corporation. 2019. MITRE ATT&CK. Retrieved September 30, 2019 from https://attack.mitre.org/.
[9] MITRE Corporation. 2019. Situation Awareness. Retrieved September 30, 2019 from https://www.mitre.org/capabilities/cybersecurity/situation-awareness.
[10] Jonathan Cryer and Kung-Sik Chan. 2008. Time Series Analysis—With Applications in R. Springer-Verlag, New York, NY.
[11] Phillip Cykana, Alta Paul, and Miranda Stern. 1996. DoD guidelines on data quality management. In Proceedings of the IQ Conference.
[12] ENISA. 2018. Exploring the Opportunities and Limitations of Current Threat Intelligence Platforms. A report by the ENISA Ad Hoc Working Group on Privacy and Technology. European Network and Information Security Agency (ENISA), Heraklion, Crete, Greece.
[13] Cosmin Ciobanu (ENISA). 2015. Standards and Tools for Exchange and Processing of Actionable Information. A report by the ENISA Ad Hoc Working Group on Privacy and Technology. European Network and Information Security Agency (ENISA), Heraklion, Crete, Greece.
[14] Center for Internet Security (CIS). 2010. CIS Security Metrics v1.1.0. Retrieved February 1, 2020 from https://www.academia.edu/3102263/CIS_Security_Metrics_v1_1_0.
[15] American Society for Quality (ASQ). 2019. Quality Assurance & Quality Control. Retrieved September 30, 2019 from https://asq.org/quality-resources/quality-assurance-vs-control.
[16] George Grispos, William Bradley Glisson, and Tim Storer. 2019. How good is your data? Investigating the quality of data generated during security incident response investigations. arxiv:1901.03723.
[17] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. 2004. Design science in information systems research. Management Information Systems Quarterly 28, 1 (2004), 75–105.
[18] SANS Internet Storm Center (ISC). 2019. InfoSec Reports. Retrieved September 30, 2019 from http://www.dshield.org.
[19] SANS Internet Storm Center (ISC). 2019. InfoSec Reports. Retrieved September 30, 2019 from https://www.dshield.org/port.html?port=3389.
[20] Wayne A. Jansen. 2009. Directions in Security Metrics Research. Technical Report NISTIR 7564. National Institute of Standards and Technology (NIST). https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7564.pdf.
[21] George Jelen. 2000. SSE-CMM security metrics. In Proceedings of the NIST & CSSPAB Workshop. https://csic.nist.rp/csspab/june13-15/jelen.pdf.
[22] Vector Guo Li, Matthew Dunn, Paul Pearce, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. 2019. Reading the tea leaves: A comparative analysis of threat intelligence. In Proceedings of the 28th USENIX Security Symposium (USENIX Security’19). 851–867. https://www.usenix.org/conference/usenixsecurity19/presentation/li.
[23] Paul W. Lowans. 2002. Implementing a Network Security Metrics Program. GIAC Administivia Version Number: 2.0. Retrieved February 1, 2020 from https://www.giac.org/paper/gsec/1641/implementing-network-security-metrics-programs/103004.
[24] Salvatore March and Gerald Smith. 1995. Design and natural science research on information technology. Decision Support Systems 15 (12 1995), 251–266. DOI: https://doi.org/10.1016/0167-9236(94)00041-2
[25] Roland Meier, Cornelia Scherrer, David Gugelmann, Vincent Lenders, and Laurent Vanbever. 2018. FeedRank: A tamper-resistant method for the ranking of cyber threat intelligence feeds. In Proceedings of the 2018 10th International Conference on Cyber Conflict (CyCon’18). 321–344. DOI: https://doi.org/10.23919/CYCON.2018.8405024
[26] Rizwan Mushqaq. 2011. Augmented Dickey Fuller test. SSRN Electronic Journal. Retrieved February 1, 2020 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1911068.
[27] Forum of Incident Response and Security Teams (FIRST). 2019. Computer Security Incident Response Team (CSIRT) Services Framework Version 2.0. Retrieved February 1, 2020 from https://www.first.org/education/csirt_services_framework_v2.0
[28] Alexandre Pinto and Alexandra Sieira. 2015. Data-Driven Threat Intelligence: Useful Methods and Measurements for Handling Indicators. Retrieved February 1, 2020 from https://www.first.org/resources/papers/conf2015/first_2015_pinto-alex_sierira-alex_data-driven-threatintelligence_20150619.pdf.
[29] Aitor Corchero Rodriguez and Mario Reyes de los Mozos. 2010. Improving network security through traffic log anomaly detection using time series analysis. In Computational Intelligence in Security for Information Systems 2010, A. Herrera, E. Corchado, C. Redondo, and A. Alonso (Eds.). Springer, Berlin, Germany, 125–133.
[30] Robert H. Shumway and David S. Stoffer. 2005. Time Series Analysis and Its Applications Springer Texts in Statistics. Springer-Verlag, Berlin, Germany.
[31] Robert H. Shumway and David S. Stoffer. 2006. Time Series Analysis and Its Applications: With R Examples (2nd ed.). Springer Texts in Statistics. Springer-Verlag, Berlin, Germany.
[32] Christian Sillaber, Clemens Sauerwein, Andrea Mussmann, and Ruth Breu. 2016. Data quality challenges and future research directions in threat intelligence sharing practice. In Proceedings of the 2016 ACM Workshop on Information Sharing and Collaborative Security (WISCS’16). 65–70. DOI: https://doi.org/10.1145/2994539.2994546
[33] Microsoft Defender ATP Research Team. 2019. Bring Your Own LOLBin: Multi-Stage, Fileless Nodersok Campaign Delivers Rare Node.js-Based Malware. Retrieved September 30, 2019 from https://www.microsoft.com/security/blog/2019/09/26/bring-your-own-lolbin-multi-stage-fileless-nodersok-campaign-delivers-rare-node-js-based-malware/?nodersok-campaign-bc-lolbins.

[34] Ray Tricker. 2001. ISO 9001: 2000 In Brief. Elsevier, San Diego, CA. https://cds.cern.ch/record/991595.

[35] Panagiotis Trimintzios. 2011. Measurement Frameworks and Metrics for Resilient Networks and Services: Challenges and Recommendations. Technical Report. European Union Agency for Cybersecurity (ENISA). https://www.enisa.europa.eu/publications/metrics-survey.

[36] Panagiotis Trimintzios. 2011. Measurement Frameworks and Metrics for Resilient Networks and Services: Technical Report. Technical Report. European Union Agency for Cybersecurity (ENISA). https://www.enisa.europa.eu/publications/metrics-tech-report.

[37] Cynthia Wagner, Alexandre Dulaunoy, Gérard Wagener, and Andras Iklody. 2016. MISP: The design and implementation of a collaborative threat intelligence sharing platform. In Proceedings of the 2016 ACM Workshop on Information Sharing and Collaborative Security (WISCS’16). ACM, New York, NY, 49–56. DOI: https://doi.org/10.1145/2994539.2994542

[38] Moira J. West-Brown, Don Stikvoort, and Klaus Peter Kossakowski. 1998. Handbook for Computer Security Incident Response Teams (CSIRTs). Technical Report. Carnegie Mellon Software Engineering Institute, Pittsburgh, PA.

[39] A. H. Yaacob, I. K. T. Tan, Su Fong Chien, and Hon Khi Tan. 2010. ARIMA based network anomaly detection. In Proceedings of the 2nd International Conference on Communication Software and Networks (ICCSN’10). 205–209. DOI: https://doi.org/10.1109/ICCSN.2010.55

[40] Z. Yong, T. Xiaobin, and X. Hongsheng. 2007. A novel approach to network security situation awareness based on multi-perspective analysis. In Proceedings of the 2007 International Conference on Computational Intelligence and Security (CIS’07). 768–772. DOI: https://doi.org/10.1109/CIS.2007.160

[41] B. Zhou, D. He, and Z. Sun. 2005. Network Traffic Modeling and Prediction with ARIMA/GARCH. Retrieved February 1, 2020 from https://www.semanticscholar.org/paper/Network-Traffic-Modeling-and-Prediction-with-ARIMA-Zhou-He/8be92869bce2c55f4d14f3f83e8e8e08427e8b5c.

[42] C. Zimmerman and C. Crowley. 2019. Practical SOC metrics. In Proceedings of the 30th Annual FIRST Conference on Computer Security Incident Handling.

[43] C. C. Zou, Weibo Gong, D. Towsley, and Lixin Gao. 2005. The monitoring and early detection of Internet worms. IEEE/ACM Transactions on Networking 13, 5 (2005), 961–974. DOI: https://doi.org/10.1109/TNET.2005.857113

Received May 2019; revised December 2019; accepted December 2019