2015 IEEE 4th Symposium on Network Cloud Computing and Applications

The Network Perspective of Cloud Security

Fabio Pierazzi, Andrea Balboni, Alessandro Guido and Mirco Marchetti

Department of Engineering “Enzo Ferrari”
University of Modena and Reggio Emilia, Italy
Email: {fabio.pierazzi,andrea.balboni,alessandro.guido,mirco.marchetti}@unimore.it

***Abstract*—The cloud computing paradigm has become really popular, and its adoption is constantly increasing. Hence, also network activities and security alerts related to cloud services are increasing and are likely to become even more relevant in the upcoming years. In this paper, we propose the first characterization of real security alerts related to cloud activities and generated by a network sensor at the edge of a large network environment over several months. Results show that the characteristics of cloud security alerts differ from those that are not related to cloud activities. Moreover, alerts related to different cloud providers exhibit peculiar and different behaviors that can be identified through temporal analyses. The methods and results proposed in this paper are useful as a basis for the design of novel algorithms for the automatic analysis of cloud security alerts, that can be aimed at forecasting, prioritization, anomaly and state-change detection.**

***Index Terms*—Security analytics; Cloud security; Cloud alerts; Temporal characterization.**

I. Introduction

In recent years, the cloud computing paradigm has known great popularity and diffusion [1]. Since all cloud services are accessed through Internet, also network activities related to their usage have been increasing as well, and this will likely become even more relevant in the upcoming years. Among all network activities related to the usage of cloud services, we focus on security alerts generated by network sensors [2]–[5], that are popular defense systems adopted for the protection of many organizations. Network sensors monitor traffic and generate a security alert whenever a packet matches a signature related to malware, botnets, scanning, or other suspicious network activity. In particular, we refer to alerts related to the usage of cloud services as *cloud security alerts*.

In this paper, we propose the first quantitative and temporal characterization of cloud security alerts observed over several months from a real large network environment. Previous works related to temporal analysis of security alerts either focus on outdated datasets (e.g., [6], [7]) or do not consider cloud activities (e.g., [8]). Our main objective is to understand if cloud alerts exhibit peculiar characteristics that could be exploited for automatic alerts analyses (e.g., aimed at anomaly detection [9]). Several endogenous and exogenous factors affect alerts generation and complicate the derivation of some conclusions, such as hosts (dis)connections, intervention of network and system administrators on firewall rules, antivirus updates. Despite all these dynamisms, the proposed characterization shows some consistent results: cloud alerts have quite different characteristics with respect to those related to non-cloud activities; alerts related to different cloud providers exhibit different temporal behaviors suggesting that they should be analyzed separately. The results and methods of our characterization are useful as a basis for the design of novel strategies for the automatic management of cloud security alerts, such as forecasting [10], anomaly [9] and state-change [11] detection. Moreover, our characterization can be useful for cloud forensics and for identifying which are the most relevant security events related to cloud activities.

The remainder of the paper is structured as follows. Section II compares the characteristics of cloud and non-cloud alerts, and motivates our investigation. Section III presents some analyses that are preliminary to the temporal characterization proposed in Section IV, where we analyze the distribution and temporal dependence of different groups of cloud alerts. Section V presents an in-depth analysis with respect to security alerts related to different cloud providers. Section VI compares our work with related literature. Finally, Section VII outlines conclusions and possible directions for future research.

II. Cloud security alerts

Our focus is on the characterization of cloud security alerts, that is, alerts generated by network activities related to cloud providers and services. In particular, we consider a real dataset of security alerts generated by a sensor at the edge of a large network environment over four months.

For this characterization, we first build a list of the major cloud providers by examining all the IaaS, PaaS and SaaS providers considered by Gartner [12] in its reports and magic quadrants related to cloud technologies and vendors. Then, we build lists of public IP addresses referring to all these cloud providers. In most cases this information is published directly in the support section of their public website (e.g., [13]). Cloud providers have an interest in making this information public and accurate since current and prospective customers can use the lists of public IPs to check whether any network malfunction is caused by firewall misconfiguration or blacklisted IPs. Whenever this information is not published, we build a list of public IP addresses based on the information that can be extracted from the RIPE public database [14]. These lists are then used to determine whether a given security alert generated by the network sensor is related to cloud activities. In particular, we classify as *cloud* each alert where the source

978-1-4673-7741-6/15 $31.00 © 2015 IEEE
DOI 10.1109/NCCA.2015.21

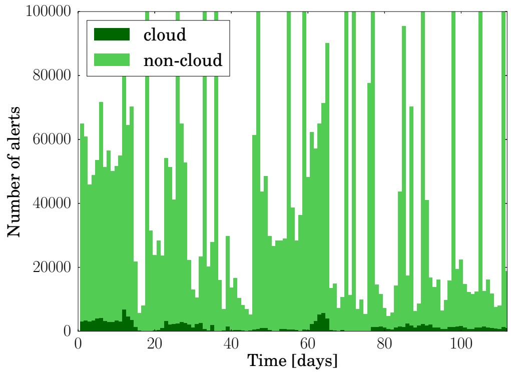


Fig. 1. Comparison of cloud and non-cloud alerts.

or destination address is included in one of the lists of public IPs related to the major cloud providers. We remark that the real dataset of security alerts has been generated by a signature-based network intrusion detection system [15] (i.e., sensor) situated at the edge of the observed large network environment. Since the sensor is used in an operational setting, it has been tuned by network administrators in order to minimize false positives. The evaluated dataset includes a total of about 160,000 cloud alerts, each related to one of the following cloud providers (listed in alphabetical order): Adobe Cloud [16], Amazon [17], CloudFlare [18], Dropbox [19], Google [20], Rackspace [21], Salesforce [22]. In Figure 1 we report a *stack histogram* of the alerts per day observed over four months, where cloud alerts numerosity is compared to non-cloud alerts. The X-axis represents time, and the Y-axis is the number of alerts per day. From this figure, we can observe that cloud alerts account for about 1.96% of the overall number of alerts generated by the network sensor, and about 98.04% of security alerts is not related to cloud activities.

Although their absolute number is low, further analyses show that cloud alerts exhibit characteristics that differ considerably with respect to non-cloud alerts. It is possible to highlight this difference by classifying security alerts with respect to the *alerts classes* defined by the taxonomy in [23]. Each class corresponds to alerts that are related to different kinds of security events and attacks (e.g., network scans, trojan activities or privilege escalation attempts). Table I shows how cloud and non-cloud alerts are distributed among the alerts classes. For each class, the column *cloud* (*non-cloud*) reports the percentage of cloud (non-cloud) alerts belonging to that class, with respect to the total number of cloud (non-cloud) alerts. As an example, we can see that 43.88% of all cloud alerts belong to the *successful-recon-limited* class (e.g., corresponding to fingerprinting or reconnaissance activities), whereas only 6.61% of all non-cloud alerts belong to the same class. In Table I we omit all the alerts classes that are defined in the taxonomy but that account for less than 0.01% for both cloud and non-cloud alerts. From this table it should be clear that the most active alerts classes are different between cloud and non-cloud alerts.

TABLE I
PERCENTAGE OF CLOUD AND NON-CLOUD ALERTS WITH RESPECT TO DIFFERENT ALERTS CLASSES.

Alerts class [23]	**Cloud**	**Non-cloud**
successful-recon-limited	43.88 %	6.61 %
web-application-attack	17.06 %	0.41 %
trojan-activity	16.36 %	73.82 %
attempted-recon	8.37 %	1.75 %
non-standard-protocol	8.01 %	0.13 %
web-application-activity	2.09 %	1.56 %
attempted-admin	1.89 %	3.26 %
misc-activity	1.54 %	3.96 %
suspicious-login	0.32 %	0.11 %
protocol-command-decode	0.23 %	0.82 %
attempted-dos	0.18 %	0.14 %
bad-unknown	0.03 %	4.06 %
attempted-user	0.02 %	0.28 %
misc-attack	0.01 %	3.06 %
suspicious-filename-detect	0.01 %	0.00 %
network-scan	0.00 %	0.03 %

Other relevant differences can be highlighted by focusing on the security alerts belonging to each class. In particular, in Table II for each alerts class we report the percentage of cloud alerts belonging to that class. Since cloud alerts are about $\approx 2\%$ of the overall number of security alerts (see Figure 1), if the distribution of cloud and non-cloud alerts was equal with respect to alerts classes, then cloud alerts would account for $\approx 2\%$ in all the classes. However, it is interesting to observe that this assumption does not hold, and that some classes contain a percentage of cloud alerts that is much higher than expected (in some cases, even higher than 10%), whereas other classes contain only few cloud alerts (less than 1%). For example, a relevant part of the alerts in *non-standard-protocol* and *web-application-attack* are related to cloud activities, whereas security alerts in *bad-unknown* and *misc-attack* are generated mostly by non-cloud activities.

These results show that the most relevant security events corresponding to cloud activities are different with respect to non-cloud alerts, thus motivating further investigations on cloud alerts characteristics and temporal behaviors. Since the growing success of cloud services, it is highly probable that such analyses will become even more relevant in the upcoming years.

III. Preliminary temporal analyses

This section presents some temporal analyses that are preparatory to the characterization proposed in the upcoming sections. We observe that any temporal analysis about the cloud security alerts is complicated by the fact that alerts generation depends on several endogenous and exogenous factors, among which we can identify:

- the number of alerts is increased by new infections, security events, attacks and attempts of attacks;

TABLE II
CONTRIBUTIONS OF CLOUD ALERTS TO EACH ALERTS CLASS.

Alerts class [23]	**Percentage of cloud alerts**
non-standard-protocol	55.49 %
web-application-attack	45.06 %
suspicious-filename-detect	21.88 %
successful-recon-limited	11.70 %
attempted-recon	8.70 %
suspicious-login	5.65 %
web-application-activity	2.61 %
attempted-dos	2.48 %
attempted-admin	1.14 %
misc-activity	0.77 %
protocol-command-decode	0.55 %
trojan-activity	0.44 %
attempted-user	0.14 %
bad-unknown	0.01 %
misc-attack	0.01 %

- the number of alerts is reduced by the manual intervention of network and cloud administrators on firewall rules, cleaning of infected machines, patching of software and operating systems, updating of antivirus and antimalware;
- finally, there are manual or automatic actions that have unpredictable effects on the number of generated alerts; for example, the number of active hosts is variable because server machines are always active, whereas clients may be disconnected during night, and novel machines can be connected to the network; updates of sensor rules may alter the signatures that generate the alerts (e.g., new IPs included in a blacklist), as well as sensor maintenance through manual shutdown or change of active ruleset.

Some of these aspects could be monitored in order to correlate them with the trend of the alerts series, whereas some of them are intrinsic to the system and likely hidden to the analysis. One of the goals of our characterization is to investigate whether, despite these dynamic factors, we can identify some peculiar temporal behaviors of the cloud alerts.

In order to reduce the impact of such noise factors, some alerts partitioning has to be performed. In particular, inspired by previous literature [24], we divide the cloud alerts in two main groups:

- *incoming cloud alerts*, that are related to packets issued from cloud services to the observed network environment;
- *outgoing cloud alerts*, that are related to packets issued from the observed network environment to cloud services.

This separation is also motivated by the fact that these two groups generate different types of alerts, as shown in Table III and Table IV (related to incoming and outgoing cloud alerts, respectively). From these tables, we can observe that most of the alerts incoming from cloud are related to *web-application-attack*, whereas most of the alerts outgoing to cloud are related to *successful-recon-limited* (e.g., attempts of gaining illegitimate access to remote services and data) and *trojan-activity* (e.g., botnets and malware activities). Hence, we can expect that different types of alerts may correspond to different temporal behaviors.

In Figure 2(a) we report a stack histogram that compares the contributions of incoming and outgoing cloud alerts, and in Figure 2(d) we report the corresponding pie chart. From these figures, we can observe that most of the activity is related to outgoing cloud alerts ($\approx 75\%$). This prevalence could be related to the fact that security policies and restrictions of cloud providers limit the number of alerts originated from cloud services. It is also interesting to investigate if there is any difference in the numerosity of the daily/nightly activity of outgoing and incoming cloud alerts. To this purpose, in Figures 2(b) and 2(c) we report the stack histograms comparing the activity during daytime hours (from 8:00 to 19:59) and night hours (from 20:00 to 7:59) of the outgoing and incoming cloud alerts, respectively. Figures 2(e) and 2(f) are the corresponding pie charts. From these figures, we can observe that most of the incoming cloud alerts are generated during daytime, whereas the outgoing cloud alerts are almost equally divided between daytime and night. A similar numerosity of cloud alerts in both daytime and night may suggest that alerts are either generated by automatic malware or originate from different time-zones [24]. On the other hand, a prevalence during daytime suggests that the alerts may be solicited by user activities, or originated from the same time-zone.

All results presented in this section motivate further investigations on the cloud security alerts by considering outgoing and incoming cloud alerts separately.

TABLE III
FIVE MOST RELEVANT CLASSES IN INCOMING CLOUD ALERTS.

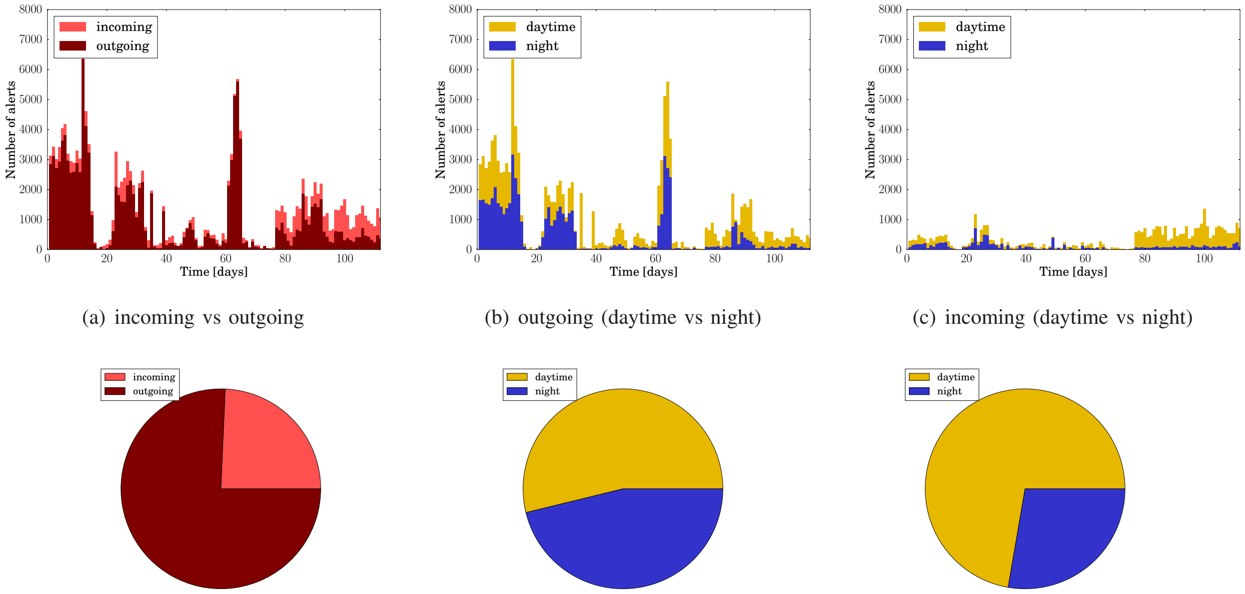
Alerts class [23]	**Percentage**
web-application-attack	67.29%
attempted-recon	8.71%
web-application-activity	8.62%
attempted-admin	7.79%
trojan-activity	4.65%

TABLE IV
FIVE MOST RELEVANT CLASSES IN OUTGOING CLOUD ALERTS.

Alerts class [23]	**Percentage**
successful-recon-limited	57.97%
trojan-activity	21.05%
non-standard-protocol	10.58%
attempted-recon	8.27%
misc-activity	2.03%

IV. TEMPORAL CHARACTERIZATION

We now characterize the temporal behavior of the outgoing and incoming cloud alerts by analyzing their distribution [25] in different time-slots of the day (Section IV-A), and by investigating the presence of temporal dependence [10] in the cloud alerts series (Section IV-B).



(a) incoming vs outgoing

(b) outgoing (daytime vs night)

(c) incoming (daytime vs night)

(d) incoming vs outgoing (pie)

(e) outgoing (pie chart)

(f) incoming (pie chart)

Fig. 2. Comparison between the incoming and outgoing cloud alerts series, and during different times of the day.

A. *Analysis of alerts distribution*

We are now interested in examining how the alerts per hour are distributed with respect to different *time-slots* of the day. We refer to hourly time-slots ranging from 0 to 23 (i.e., from 12am to 11pm). We consider a time granularity equal to one hour because it guarantees an acceptable compromise between noises and patterns for alerts generated in large network environments [26]. The purpose is to examine how the numbers of cloud alerts per hour are distributed with respect to daily and nightly time-slots. This is useful for increasing the awareness on the security status of the system, and is also a preparatory step for forecasting, prioritization, anomaly and state-change detections [9], [11], [27] of cloud alerts tuned on the observed environment.

The alerts distribution analysis begins by considering the cloud alerts series per hour C_t, where each element c_t represents the number of cloud alerts generated during hour t. In particular, we consider two separate series related to the outgoing and incoming cloud alerts, respectively. In order to evaluate the cloud alerts distribution and dispersion, each series C_t is sampled into 24 separate datasets $d_0, d_1, ..., d_{23}$, where each dataset d_i contains the number of cloud alerts per hour detected during the i-th time-slot of the day ($i \in \{0, 1, ..., 23\}$). For example, the dataset d_2 contains the values of the number of cloud alerts per hour that were generated from 2:00am to 2:59am.

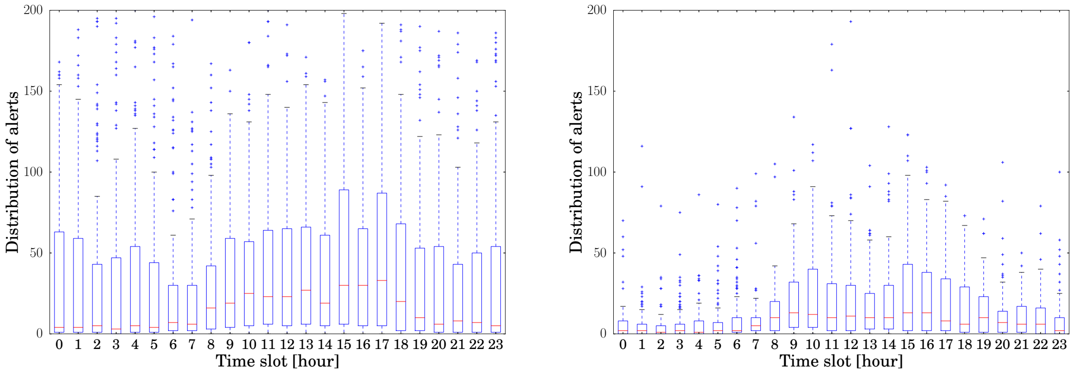
In Figures 3, we report side-by-side boxplots [25] referring to the time-slots of the outgoing and incoming cloud alerts, respectively. This representation of the alerts distribution is useful because it easily allows us to:

- understand how alerts are distributed and dispersed with respect to the time-slots of the day;
- estimate number and scale of outliers with respect to the different time-slots;
- compare behavior of outgoing and incoming cloud alerts.

Another important aspect of this representation is that it contains simultaneously information about the cloud alerts *hourly* activity (because each boxplot represents the distribution of the number of alerts per hour), and information about the temporal behavior of the cloud alerts during the *day* (because the boxplots of the different time-slots are represented side-by-side).

Let us first focus on the distribution of outgoing cloud alerts in Figure 3(a). We can observe that the alerts dispersion (represented by interquartile ranges) is similar among the different time-slots of the day. This is probably related to automatic activities that are executed by bots or other malware installed on hosts that are always active. It is interesting to observe that, however, the medians are more active during daily time-slots between 8 and 18. This suggests that part of the outgoing cloud alerts may be related to user activities (e.g., hosts turned on/off by users, or user interactions with cloud services). On the other hand, in Figure 3(b) we can observe that both dispersion and medians of incoming cloud alerts are higher during daily time-slots and lower during night, thus suggesting that most of incoming cloud alerts activity may be caused in response to users interactions.

These results show how our analyses are effective for understanding relevant temporal information about cloud alerts distribution, and for acquiring more awareness on the security status of the system.



(a) outgoing cloud alerts

(b) incoming cloud alerts

Fig. 3. Analysis of alerts distribution for outgoing and incoming cloud alerts.

B. Analysis of temporal dependence

We now investigate the presence of temporal dependence in the outgoing and incoming cloud alerts by analyzing if their series exhibit some relevant trend, periodic or seasonal components, or if they are dominated by noise. The trend represents a systematic component of the series that does not repeat over time, whereas periodic and seasonal components repeat within the time range captured by the data; noise represents a component that could hide trends and periodicities. The analysis of temporal dependence is also useful as a basis to determine whether cloud alerts series are predictable or not [10], and if anomaly or state-change detection approaches could be applied effectively [9], [11], [27].

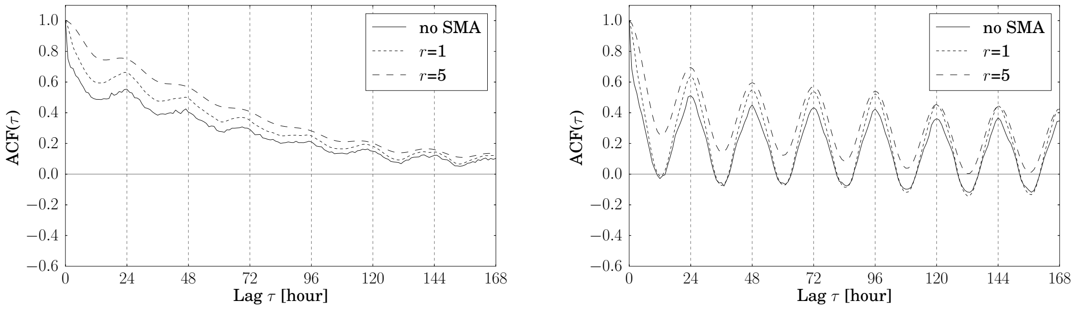
In order to investigate the presence of temporal dependence, we consider the *autocorrelation function* [10], and we define $ACF(\tau)$ as the value of the autocorrelation function at lag τ. High values and slow decay of the ACF suggest that future values are related to past values with some degree of accuracy. In particular, a time series is considered predictable for a window k if its autocorrelation function $|ACF(i)| \geq 0.3, \forall i \in \{0, 1, ..., k\}$ [10], [28]. Moreover, autocorrelation studies can also reveal the presence of periodicity in the analyzed series.

Since temporal dependence might be hidden by noise and/or out-of-scale outliers, before evaluating the ACF we also perform some filtering on the cloud alerts series. First, we replace the outliers above the 99th quantile with the value of the upper whisker, because out-of-scale values could corrupt the autocorrelation analyses. Then, since we are not interested in finding the optimal filtering technique for each cloud series, we adopt a simple smoothing filter that does not alter the nature of the data. In particular, we consider a *simple moving average* (SMA) filter with a centered window of radius r hours, where each value of the series C_t is replaced with the average of its $2r$ neighbors. In order to evaluate if the autocorrelation results are influenced by filtering, we consider three different configurations: no filtering, SMA filter with radius $r = 1$ (3-hour window) for a low-impact smoothing, and SMA filter with radius $r = 5$ (11-hour window) for a stronger smoothing.

In Figures 4 we report the autocorrelation results related to the outgoing and incoming cloud alerts, respectively. In particular, each figure reports results related to the three configurations: no filtering, SMA filtering with radius $r = 1$ and $r = 5$. The X-axis represents the autocorrelation lag τ in hours, and the Y-axis reports the ACF values. The vertical dashed lines represent 24-hour shifts. From Figure 4(a), we can observe that the outgoing cloud alerts exhibit a strong trend component with a slow decay, and that the series is predictable for even 72 hours ahead. This result may also be related to the similar dispersion of the outgoing alerts in the different time-slots of the day (see Figure 3(a)). On the other hand, in Figure 4(b) the incoming cloud alerts exhibit a strong 24-hour periodicity, thus implying that the highest probability of finding a similar value is 24 hours ahead. The information about this periodicity can be useful for modeling the series for prediction and anomaly detection purposes [9], [10]. Moreover, this 24-hour periodicity suggests that most of the incoming cloud alerts are probably related to user interactions.

V. Cloud providers characterization

In this section, we perform a further breakdown of the outgoing and incoming cloud alerts by considering the three most active providers in the observed environment. The purpose is to investigate whether alerts related to different cloud providers exhibit different temporal behaviors. For the sake of fairness, we refer to the three most active providers as *CP1*, *CP2* and *CP3*, in descending order of numerosity of alerts. In Figures 5, we report the stack histograms and pie charts that compare outgoing and incoming alerts of CP1, CP2 and CP3, respectively. We can observe that most of the outgoing cloud alerts are related to CP1, whereas most of the incoming cloud alerts are related to CP2. An interesting observation is that CP1 and CP2 are used to deliver services targeted to the users in the observed network environment. In particular, CP1 is used mainly as a SaaS provider, hence it is plausible that the number of outgoing cloud alerts is higher, since the alerts are mainly



(a) outgoing cloud alerts

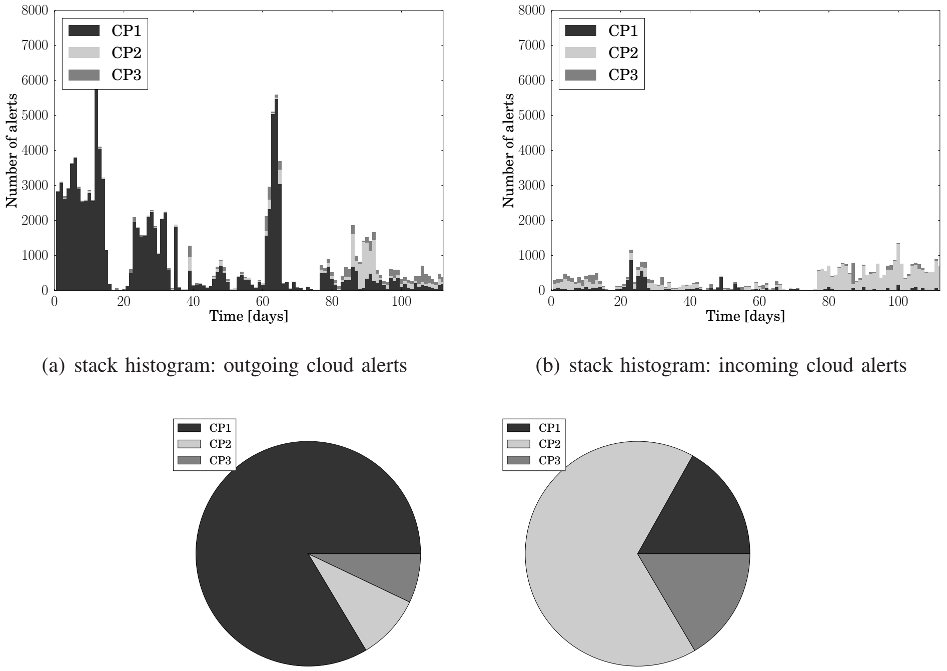
(b) incoming cloud alerts

Fig. 4. Analysis of temporal dependence for outgoing and incoming cloud alerts.

related to internal hosts that contact providers machines. On the other hand, CP2 has been adopted from around day 75 (see Figure 5(b)), and is used mainly for content delivery and proxying, hence it is plausible that most alerts are related to cloud activities coming from outside the observed network (incoming cloud alerts). Finally, the alerts of CP3 are lower in terms of numerosity also because CP3 is not used directly for offering services to the users in the observed network environment.

This different behavior of the cloud providers is relevant when considering automatic algorithms for the management of huge volumes of alerts. For example, the incoming alerts of CP2 in Figure 5(b) exhibit a rather stable behavior in terms of number of cloud alerts per day, with a state-change around day 75, hence even simple threshold-based algorithms (e.g., CUSUM-based [11]) considering a time granularity equal to a day could be effective for identifying relevant anomalies and/or state-changes in this alerts group. On the other hand, in Figure 5(a) we can observe that the cloud alerts series per day of CP1 is more unstable, hence considering finer time granularities may be more appropriate when modeling algorithms for the detection of relevant security events in this group.

In Figures 6, we present a more detailed analysis of cloud alerts distribution and temporal dependence with respect to a finer time granularity of an hour. For space reasons, we report the most relevant results referring to CP1 (outgoing), CP2 (incoming) and CP3 (incoming). These figures confirm



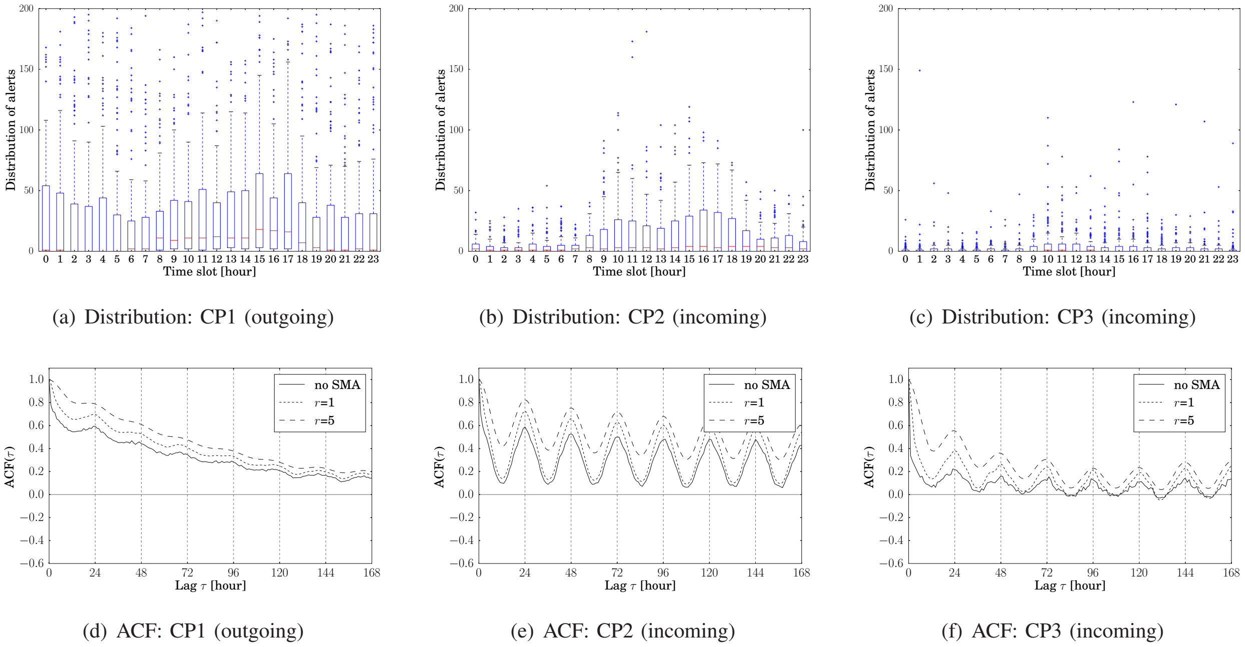
(a) stack histogram: outgoing cloud alerts

(b) stack histogram: incoming cloud alerts

(c) pie chart: outgoing cloud alerts

(d) pie chart: incoming cloud alerts

Fig. 5. Contribution of different providers with respect to incoming and outgoing cloud alerts.



(a) Distribution: CP1 (outgoing)

(b) Distribution: CP2 (incoming)

(c) Distribution: CP3 (incoming)

(d) ACF: CP1 (outgoing)

(e) ACF: CP2 (incoming)

(f) ACF: CP3 (incoming)

Fig. 6. Comparison of alerts distribution and temporal dependence related to different cloud providers.

that most of the behavior of outgoing and incoming cloud alerts is mainly caused by CP1 and CP2, respectively (see also Figures 3). In particular, we have that the dispersion of CP1 (outgoing) is similar with respect to the different time-slots (Figure 6(a)), and the high autocorrelation suggests that the trend of this series is predictable for even 72 hours ahead (Figure 6(d)). Most of the activity of CP2 (incoming) is focused during daytime (Figure 6(b)), with a strong 24-hour periodicity (Figure 6(e)). On the other hand, the number of cloud alerts related to CP3 (incoming) is lower (Figure 6(c)), with a slight prevalence of activity during daily time-slots. Although Figure 6(f) shows that the autocorrelation of CP3 (incoming) is lower with respect to the other cloud providers, we can observe that the smoothing filter with radius $r = 5$ improves predictability and highlights a weak 24-hour periodicity.

The results presented in this section show that a separation based on the cloud providers can be effective for a more accurate modeling of the cloud alerts characteristics, that is preparatory for automatic analyses.

VI. Related Work

To the best of our knowledge, this paper proposes the first quantitative and temporal characterization of cloud security alerts observed from a real large network environment over several months. These results are useful as a basis for further studies on automatic analyses of cloud security alerts, that are likely to become more relevant in the upcoming years.

The work proposed in this paper mainly relates to three research areas: attribute-based alerts correlation, temporal analysis of security alerts, and cloud security.

Most of the previous work focused on security alerts propose some sort of correlation algorithms, mostly based on alerts attributes [29], [30]. Their main goal is to aggregate alerts having similar attributes, such as source/destination addresses or timestamps. For example, normalization and fusion unify alerts coming from different sources (e.g., through IDMEF format [31]); prioritization [32] associates a level of risk to each alert on the basis of an asset database; verification [33] determines through heuristics whether an attempt of attack has been successful or not (e.g., an alert for a Windows vulnerability directed at a Linux server); multi-step attack detection [34], [35] aims at identifying alerts that are part of the same attack. Although the approach proposed in this paper focuses on alerts analysis, it clearly differs from these previous works. Most approaches based on attribute-based alerts correlation work well for relatively stable contexts and/or require a-priori knowledge on attack scenarios and on the characteristics of the monitored environment. On the other hand, characterizations and temporal analyses proposed in this paper aim at identifying peculiar temporal characteristics and properties that can be useful for designing novel algorithms and strategies for the management of cloud security alerts.

Temporal analysis of security alerts is considered by Qin et al. [6] for correlating alerts series to identify novel attacks. However, their work has a different goal (multi-step attack detection vs preliminary temporal analyses), relies on different techniques, and performs its evaluation on the outdated DARPA dataset [7]. On the other hand, our focus is on characterizing quantitative and temporal properties of cloud security alerts and our analyses refer to a dataset that comprises real and recent security alerts. Other works related to temporal analysis of security alerts are presented in some papers by

Viinikka et al. [8], [26], [36], where the authors propose several techniques for anomaly detection of low-priority alerts series (e.g., related to ICMP messages). However, they assume that data exhibit strong temporal dependence, and focus on the proposal of anomaly detection algorithms, whereas our focus is on the characterization of cloud security alerts, that can also be useful as a basis for understanding applicability of regression-based anomaly detection [9].

Finally, the work proposed in this paper clearly differentiates from the existing literature related to cloud security [37], that mainly focuses on the proposals of architectures for guaranteeing data (e.g., [38], [39]) and network (e.g., [40]) protection in cloud platforms.

VII. Conclusions

In this paper, we propose the first quantitative and temporal characterization of security alerts related to cloud network activities. This investigation is preliminary for identifying and tuning the most appropriate techniques for the automatic analysis of cloud security alerts, such as forecasting, prioritization, anomaly and state-change detection. Results referring to real cloud alerts generated by a network sensor at the edge of a large network environment show that our analyses are able to identify different characteristics between cloud and non-cloud security alerts. Moreover, in-depth analyses on cloud alerts show that alerts related to different cloud providers exhibit different temporal behaviors, hence suggesting that they should be studied and modeled separately. These results have been achieved despite the dynamism and several noise factors inherent to the analyzed context. Future work will focus on the study of methodologies for the automatic investigation of cloud alerts characteristics with the purpose of understanding applicability of popular alerts management techniques.

References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[2] B. Mukherjee, L. T. Heberlein, and K. N. Levitt, "Network intrusion detection," *IEEE Network*, vol. 8, no. 3, pp. 26–41, 1994.

[3] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, no. 2, pp. 222–232, 1987.

[4] M. Roesch, "Snort: Lightweight intrusion detection for networks." in *Proc. of the 13th USENIX Conference on System Administration (LISA-99)*, vol. 99, 1999, pp. 229–238.

[5] M. Colajanni and M. Marchetti, "A parallel architecture for stateful intrusion detection in high traffic networks," in *Proc. of the IEEE/IST Workshop on Monitoring, Attack detection and Mitigation (MonAM)*, Sept. 2006.

[6] X. Qin and W. Lee, "Statistical causality analysis of infosec alert data," in *Proc. of the 6th International Symposium on Recent Advances in Intrusion Detection*. Springer, 2003, pp. 73–93.

[7] R. P. Lippmann, D. J. Fried, I. Graf, J. W. Haines, K. R. Kendall, D. McClung, D. Weber, S. E. Webster, D. Wyschogrod, R. K. Cunningham *et al.*, "Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation," in *Proc. DARPA Information Survivability Conference and Exposition*, vol. 2, 2000, pp. 12–26.

[8] J. Viinikka, H. Debar, L. Mé, A. Lehikoinen, and M. Tarvainen, "Processing intrusion detection alert aggregates with time series modeling," *Information Fusion*, vol. 10, no. 4, pp. 312–324, 2009.

[9] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, p. 15, 2009.

[10] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*. Taylor & Francis, 2002.

[11] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*. Prentice Hall Englewood Cliffs, 1993.

[12] "Gartner," http://www.gartner.com/, visited in May 2015.

[13] "Amazon Web Services: IP ranges," http://docs.aws.amazon.com/general/latest/gr/aws-ip-ranges.html, visited in May 2015.

[14] "Ripe Network Coordination Centre," https://stat.ripe.net/, visited in May 2015.

[15] "Suricata IDS," http://suricata-ids.org/, visited in May 2015.

[16] "Adobe Cloud," https://www.adobe.com/creativecloud.html, visited in May 2015.

[17] "Amazon Web Services," http://aws.amazon.com/, visited in May 2015.

[18] "CloudFlare," https://www.cloudflare.com/, visited in May 2015.

[19] "Dropbox," https://www.dropbox.com/, visited in May 2015.

[20] "Google cloud platform," https://cloud.google.com/, visited in May 2015.

[21] "Rackspace," http://www.rackspace.com/, visited in May 2015.

[22] "Salesforce," http://www.salesforce.com/, visited in May 2015.

[23] "Snort users manual," http://manual.snort.org/, visited in May 2015.

[24] D. Dagon, C. C. Zou, and W. Lee, "Modeling botnet propagation using time zones." in *Proc. of the 13th Network and Distributed System Security Symposium*, vol. 6, 2006, pp. 2–13.

[25] T. T. Soong, *Fundamentals of probability and statistics for engineers*. John Wiley & Sons, 2004.

[26] J. Viinikka, H. Debar, L. Mé, and R. Séguier, "Time series modeling for IDS alert management," in *Proc. of the ACM Symposium on Information, Computer and Communications Security (AsiaCCS)*, 2006, pp. 102–113.

[27] D. C. Montgomery, *Introduction to statistical quality control*. John Wiley & Sons, 2007.

[28] S. Casolari, S. Tosi, and F. Lo Presti, "An adaptive model for online detection of relevant state changes in internet-based systems," *Performance Evaluation*, vol. 69, no. 5, pp. 206–226, 2012.

[29] F. Valeur, G. Vigna, C. Kruegel, and R. A. Kemmerer, "Comprehensive approach to intrusion detection alert correlation," *IEEE Transactions on Dependable and Secure Computing*, vol. 1, no. 3, pp. 146–169, 2004.

[30] H. Debar and A. Wespi, "Aggregation and correlation of intrusion-detection alerts," in *Proc. of the 4th International Symposium on Recent Advances in Intrusion Detection*. Springer, 2001, pp. 85–103.

[31] H. Debar, D. A. Curry, and B. S. Feinstein, "The intrusion detection message exchange format (IDMEF)." 2007.

[32] M. Colajanni, M. Marchetti, and M. Messori, "Selective and early threat detection in large networked systems," in *Proc. of the 10th IEEE International Conference on Computer and Information Technology (CIT)*, Bradford, UK, June 2010.

[33] C. Kruegel and W. K. Robertson, "Alert verification determining the success of intrusion attempts," in *Proc. First Workshop the Detection of Intrusions and Malware and Vulnerability Assessment (DIMVA)*, 2004, pp. 25–38.

[34] S. T. Eckmann, G. Vigna, and R. A. Kemmerer, "STATL: An attack language for state-based intrusion detection," *Journal of Computer Security*, vol. 10, no. 1, pp. 71–103, 2002.

[35] M. Marchetti, M. Colajanni, and F. Manganiello, "Framework and Models for Multistep Attack Detection," *International Journal of Security and Its Applications*, vol. 5, no. 4, pp. 73–92, Oct. 2011.

[36] J. Viinikka and H. Debar, "Monitoring IDS background noise using EWMA control charts and alert information," in *Proc. of the 7th International Symposium on Recent Advances in Intrusion Detection*. Springer, 2004, pp. 166–187.

[37] T. Mather, S. Kumaraswamy, and S. Latif, *Cloud security and privacy: an enterprise perspective on risks and compliance*. O'Reilly Media, Inc., 2009.

[38] R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan, "CryptDB: protecting confidentiality with encrypted query processing," in *Proc. 23rd ACM Symp. Operating Systems Principles*, Oct. 2011.

[39] L. Ferretti, F. Pierazzi, M. Colajanni, and M. Marchetti, "Scalable architecture for multi-user encrypted SQL operations on cloud database services," *IEEE Transactions on Cloud Computing (TCC)*, vol. 2, no. 4, pp. 448–458, Oct. 2014.

[40] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in cloud," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 42–57, 2013.