\textbf{PERMON}: An OpenStack Middleware for Runtime Security Policy Enforcement in Clouds

Azadeh Tabiban*, Suryadipta Majumdar*, Lingyu Wang* and Mourad Debbabi*

*Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

Email: \{a_tabiba, su_majum, wang, debbabi\}@ciise.concordia.ca

\section*{Abstract}—To ensure the accountability of a cloud environment, security policies may be provided as a set of properties to be enforced by cloud providers. However, due to the sheer size of clouds, it can be challenging to provide timely responses to all the requests coming from cloud users at runtime. In this paper, we design and implement a middleware, \textit{PERMON}, as a pluggable interface to OpenStack for intercepting and verifying the legitimacy of user requests at runtime, while leveraging our previous work on proactive security verification to improve the efficiency. We describe detailed implementation of the middleware and demonstrate its usefulness through a use case.

\textbf{Keywords}—Access Control, Event Interception, Middleware, Cloud Security, OpenStack.

\section{I. Introduction}

The multi-tenancy nature of clouds has proved to be a double-edged sword, since it leads to the main advantage of resource optimization while at the same time causes many inherent security concerns \cite{19}. On the other hand, the self-service nature coupled with the sheer size of the cloud renders it a challenging task to enforce tenants’ security policies at runtime. In fact, verifying every user event at runtime can cause considerable delays (e.g., over four minutes \cite{3}) even in a mid-sized cloud.

To that end, a promising solution for reducing the response time of security policy enforcement in clouds to a practical level is the proactive approach (e.g., \cite{3, 10, 11}). Such an approach prepares for expensive verification tasks in advance, based on probable temporal relationships between user requests (e.g., a user is likely to create security groups after creating a new VM in OpenStack \cite{15}). Our previous works \cite{10, 11} have demonstrated that the proactive verification approach may process user requests with only negligible delays.

In this paper, we apply the proactive approach to OpenStack \cite{15}, which is one of the most popular cloud platforms, through designing and implementing a middleware, namely \textit{PERMON}. The middleware is designed to first intercept user-issued requests on their path to an intended service. It then identifies the requested event types based on either a single or an aggregation of differentiating fields of corresponding requests. Finally, the middleware processes selected parameters coupled with the identified event types, and enforces the verification results by either granting or rejecting the user requests. Specifically, our main contributions are as follows.

- We apply our proactive verification approach to OpenStack as an efficient middleware solution for enforcing security policies at runtime in clouds.
- We detail the implementation of our approach as a Web Server Gateway Interface (WSGI) pluggable middleware for OpenStack.
- We demonstrate the usefulness of our middleware through describing a use case.
- We address the key challenge of event type identification and discuss the additional benefit of our solution to log processing.

The remainder of the paper is organized as follows. Section II provides background and related work. Section III gives our middleware design. Section IV provides the implementation details. Section V describes a use case. Section VI includes further discussions. Section VII concludes the paper.

\section{II. Preliminaries}

In this section, we first provide a quick review of our proactive verification approach \cite{10, 11} to facilitate further discussions, and then review related work.

\subsection{A. Proactive Security Verification}

Our middleware leverages our previous proactive verification approach, LeaPS \cite{10, 11}. The main idea of LeaPS is to prepare for expensive verification tasks in advance, before actual events are received. Such preparation is possible due to probable temporal relationships between user events, e.g., creating a VM and creating security groups for the same VM, which can be automatically extracted from event logs using machine learning techniques \cite{11}. The results of such preparation (i.e., verification results for possible next events) can then be stored in a table, such that the actual verification performed after events are received amounts to a simple lookup inside the table, which takes far less time than the actual verification does.

To illustrate the idea, Figure 1 compares how user requests are processed under a traditional intercept-and-check approach and under our proactive solution, respectively. In the upper timeline, an intercept-and-check approach intercepts and then verifies the update port user request against the desired security property “no bypass” for the anti-spoofing mechanisms in the cloud, which can be violated by real world vulnerabilities, e.g., OSSA-2015-018 \cite{13}. A traditional intercept-and-check approach (e.g., \cite{3}) would take up to several minutes in medium
size clouds to determine whether the request should be granted or denied, which is apparently unacceptable to the user who has issued the request.

In contrast, as depicted in the lower timeline, the proactive LeaPS approach works very differently: It proactively conducts a set of pre-computations distributed among $N$-steps ahead of the actual occurrence of the critical operation (Update port). The pre-computation is based on a dependency model which captures the probable temporal order between user events. The dependency model may be manually created based on structural dependencies inherent to the cloud platform [10], e.g., a security group can only be created after the VM is created. The model can also be automatically established through applying machine learning techniques to extract frequent patterns or sequences of events from the logs, which may correspond to not only aforementioned cloud platform-specific structural dependencies but also other runtime dependencies, e.g., those due to business rules or user habits [11].

The pre-computations incrementally prepare the needed information for efficiently verifying the critical operation later on. Specifically, the verification results of potential future events are pre-computed and stored in a so-called watchlist, and consequently the actual verification only takes negligible time (e.g., six milliseconds [10]) after the actual events are received. In this paper, we design and implement the middleware, PERMON, to leverage LeaPS as its proactive verification engine. We will detail the challenges faced in designing PERMON to work with LeaPS and provide solutions in later sections.

B. Related Work

There exist many solutions for enforcing security compliance in the cloud. For instance, Solonas et al. [20] propose an approach to detect illegal and undesired activities in the cloud based only on collected billing data in order to preserve privacy. In [12], [9], formal audit approaches have been proposed for security compliance checking in the cloud. Unlike our middleware, those approaches can detect violations only after the fact, which may expose the system to unrecoverable damages. VeriFlow [7] and NetPlumber [6] monitor network events and check network properties and policies at runtime to capture bugs before or as soon as they occur. They rely on incremental calculations to achieve the runtime verification. These works focus on operational network properties (e.g. black holes and forwarding loops) in traditional networks, whereas our effort is oriented toward cloud virtualized infrastructures.

In the context of runtime security monitoring in the cloud, various mechanisms and concepts for designing security service level agreement-based cloud monitoring services have been discussed in [17]. CloudSec [5] and CloudMonatt [25] are solutions that have been proposed for VMs security monitoring. In contrast, our work can potentially cover generic security properties beyond the scope of VMs. In addition, rather than intercepting security measurements, we intercept actual events and assess their impact on the cloud system before applying them. In [16], a host-based secure active monitoring mechanism has been proposed. Unlike our work which focuses on high level policies, the main objective of this work is to detect unwanted low level operations initiated by malicious software.

Retroactive auditing approaches (e.g., [9], [12], [22], [23], [21], [4]) in the cloud is a traditional manner to verify the compliance of different components of a cloud. In [9], [12], formal auditing approaches are proposed for retroactive security compliance checking in the cloud. The intercept-and-check approach performs major verification tasks while holding the event instances blocked. Weatherman [3] and OpenStack Congress [14] offer security verification of virtual infrastructure using the intercept-and-check approach, and causes significant delay to a user request. Unlike those works, we provide a proactive enforcement approach which enables starting the verification process in advance.

Proactive security analysis in the cloud is comparatively a new domain with fewer works (e.g., [3], [18], [24], [10], [11]). Weatherman [3] verifies security policies on a future change plan in a virtualized infrastructure using the graph-based model proposed in [2], [1]. Our previous work, PVSC [10], proactively verifies security compliance by utilizing the static patterns in dependency models. Both in Weatherman and PVSC, models are captured manually by expert knowledge. In contrast, this work is an extension of our previous work LeaPS [11], which adopts a learning-based approach to automatically derive the dependency model. Closest to our work, the OpenStack Security Modules (OSM) [8] is an access control framework for OpenStack to support different access control modules by replacing the existing permission checks in OpenStack; the framework includes a service called patron which intercepts user requests similarly as our middleware, whereas the policy enforcement is not proactive like ours.
III. The Design

In this section, we provide the high level design of PERMON, a proactive security policy enforcement middleware for OpenStack. The main objective of the middleware is to intercept and verify each user request against given security policies and properties such that the request is either denied or allowed to reach the intended service. We will leverage our proactive verification component LeaPS (see Section II-A), although PERMON is designed to work with any other solution which can perform security verification. Figure 2 provides a high level overview about how PERMON works. The process consists of four main steps: Users’ requests are passed through PERMON on their paths to the intended services (step 1). PERMON intercepts the requests and extracts parameters which are pre-determined according to the mechanism of the deployed security solution (the proactive verification component LeaPS in our case) (step 2). These parameters are verified by the security solution (LeaPS) (step 3). The verification result on the legitimacy of the requested action is put into effect by PERMON (step 4).

![Fig. 2. An overview of PERMON](image)

The proactive aspect of PERMON works as follows. The main objective is to pre-compute the conditions to be verified at the moment when the so-called critical events (i.e., the events that may directly cause breaches to given security properties) are received. The critical events and those events that are likely to precede the critical events have been pre-defined in the dependency model in advance. At runtime, once PERMON intercepts any of those latter events, the verification module will be triggered to precompute the required conditions determined by the imposed security property. When the actual critical event is intercepted by PERMON, those precomputed conditions are verified through simple table lookups by analyzing the intercepted attributes, and the legitimacy of the critical event is decided and enforced by PERMON.

The main challenges of implementing this design of PERMON includes intercepting user requests, identifying types of events, interacting with the verification module, and enforcing the received verification results, which will be detailed in next section.

IV. Implementation

In this section, we detail PERMON implementation and its integration to OpenStack. Figure 3 shows the architecture of PERMON with examples of inputs to various steps performed by PERMON. Algorithm IV describes those steps in more details. In the following, we elaborate on the implementation and our approach of employing the general mechanism of a Web Server Gateway Interface (WSGI) for implementing the middleware functionality.

**Algorithm 1 PERMON (APIcalls, mark-rules, field-check-rules)**

```plaintext
1: procedure POLICYENFORCEMENT(APIcalls)
2:   for each call_i in APIcalls do
3:     env[] = readingEnviron(call_i)
4:     eventToWatch = markFinder(env[], mark-rules)
5:     if eventToWatch is NULL then
6:       response = callNextComponent(call_i)
7:     else
8:       fields = readFields(env, field-check-rules)
9:       Decision = LeaPS(fields, eventToWatch)
10:      if Decision is Allow then
11:         response = callNextComponent(call_i)
12:      else
13:         response = callFailed("403 Forbidden")
14:      return response
```

A. Implementing PERMON as a WSGI Middleware

Many features of OpenStack have been implemented as pluggable middlewares based on the Web Server Gateway Interface (WSGI) standard. The main advantage to this approach is that the pluggable nature of deployed middlewares provides developers with flexibility to extend existing functionality without undertaking the cost of customizing existing codes. Following this approach, PERMON is designed as a WSGI interface, and is injected into Nova (OpenStack compute service) pipeline alongside other middlewares being stacked up.

Upon receiving clients’ requests, each middleware in the pipeline takes the request sent from the previous component, and calls its next adjacent middleware, which means each middleware acts both as a server and an application. Any element that can play the role of a WSGI application has a call method that is passed with two positional arguments when it is executed by the preceding server; the first is a python dictionary for environment information, called environ, and the second is a callback function named start_response, as detailed below.

- **environ** is a python dictionary that contains contextual information of a request. Such information includes: REQUEST_METHOD, PATH_INFO, wsgi.input, etc. Examples of the REQUEST_METHOD include PUT, GET, POST, etc. PATH_INFO is the URL path an API request is sent to, and wsgi.input is a file-like object and contains data that is sent to the server.
- **start_response** is a callable itself, and takes two positional arguments, status and header; status is an integer followed by a string, and header is a list of tuples that are sent back to the client sending the request.
Having been inserted into the pipeline, PERMON is invoked by its preceding element for each request made to Nova, which is passed with the environ and start_response arguments corresponding to the request. PERMON here serves the role of a server for its next element in the pipeline. Inspecting environ, we can find useful information to process the corresponding request as we will explain in more details in next section.

### B. Event Type Identification

A key challenge in implementing PERMON is to identify the intended actions in users’ requests intercepted by PERMON. The event type corresponding to such actions will determine which actions to be taken while interacting with the proactive verification module. By analyzing OpenStack API documentation, some API requests can be uniquely identified through the conjunction of their METHOD and PATH_INFO attributes. For instance, a given request can be recognized as to delete an instance if and only if its corresponding METHOD and PATH_INFO is DELETE and `/servers/server_id` respectively.

<table>
<thead>
<tr>
<th>INPUT wsgi.input Content</th>
<th>METHOD</th>
<th>PATH_INFO</th>
<th>Event Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>{&quot;os-stop&quot;: null}</td>
<td>POST</td>
<td>/servers/server_id/action</td>
<td>Stop VM</td>
</tr>
<tr>
<td>{&quot;os-start&quot;: null}</td>
<td>POST</td>
<td>/servers/server_id/action</td>
<td>Start VM</td>
</tr>
</tbody>
</table>

However, not all API requests can be uniquely identified through the conjunction of their METHOD and PATH_INFO attributes. For example, as it is shown in Table I, the issued requests for starting and stopping a given server both have the same METHOD and PATH_INFO environmental variables according to API documentation. Therefore, we need another differentiating attribute to fingerprint this group of requests. By studying API documentation, we find that the METHOD field of all these requests is POST. As the data is sent in wsgi.input file-like environmental variable for requests with PUT and POST as their METHOD, we use its content as the differentiating attribute. For example, the content of the wsgi.input object for the two requests (for starting and stopping a given server) is {"os-start": null} and {"os-stop": null}, respectively (Table I), which is intercepted and parsed by PERMON as the differentiating attribute.

By parsing the body of requests together with the combination of METHOD and PATH_INFO attributes, PERMON can map them to the type of event they are issued for. This is performed through knowing the collection of differentiating attributes of a given request and a rule mapping these attributes to a pre-defined type of event. To that end, we have embedded inside PERMON a set of mapping rules which associate the name of the differentiating attributes and their content with the corresponding event types. For example, the wsgi.input attribute for attaching a VM to security group SG1 is {"addSecurityGroup": {"name":"SG1"}}. To investigate whether a received request is issued to add a VM to a security group, PERMON examines wsgi.input attribute, and verifies its starting characters.

As mentioned earlier, we leverage our proactive verification approach, LeaPS, which triggers pre-computation at the occurrence of certain events that are determined to be more likely to be followed by a critical event. Therefore, our mapping rules for event type identification are designed to map the request attributes to those two types of events. When the former type of events are intercepted, PERMON will invoke LeaPS for pre-
computation, and when the latter type of events are intercepted, \textit{PERMON} will invoke \textit{LeaPS} for performing the verification.

\subsection{Interacting with \textit{LeaPS} and Enforcing the Decisions}

Depending on the event types, \textit{PERMON} will invoke \textit{LeaPS} either for pre-computation or for verification. During such interaction, in addition to the event types, other parameters of the corresponding event need to be extracted and passed to \textit{LeaPS} in order to verify the legitimacy of the request or to pre-compute conditions. The exact parameters needed are determined based on the security properties and event types. For example, verifying requests against a security property may require the security group name and the VM ID for the event type \textit{add security group}. Therefore, we have implemented inside \textit{PERMON} the mappings between event types and rules to extract required parameters from the intercepted environmental variables of the corresponding request. More specifically, the extraction rules will determine which attributes to be fed into \textit{LeaPS}, and the way they should be pre-processed by \textit{PERMON}. Finally, \textit{PERMON} enforces the verification results of \textit{LeaPS} over the legitimacy of intercepted requests. If \textit{LeaPS} allows the event, \textit{PERMON} calls the next component in the pipeline and passes to it the intercepted arguments for further processing. If \textit{LeaPS} indicates denial of the request, \textit{PERMON} blocks the request by controlling the \texttt{start_response} function as follows. One of the positional arguments, status, is fed into the callback function \texttt{start_response}. The value of this status argument of a request that has passed successfully through the previous component in the pipeline will start with a special value \texttt{2xx}. By changing this argument value to \texttt{403 Forbidden}, \textit{PERMON} essentially blocks the request and an error message will be sent back to the user.

\section{Use Case}

In this section, we demonstrate the usefulness of \textit{PERMON} through an example use case. We assume users’ requests are made from OpenStack console. The following illustrates the functionalities of \textit{PERMON}, which is to intercept the requests and their parameters, to interact with the verification module \textit{LeaPS}, and to allow or block the corresponding requests according to the verification result.

The security property to be enforced in the use case is no \textit{downgrade of security group for a running VM}, which prevents attackers from changing the security group of a VM, e.g., from \textit{no_connection} to \textit{essential} (the latter is supposedly a less restricted security group which will allow more connections to the VM, downgrading its security level). For each security group, the verification module \textit{LeaPS} is initialized with security groups with higher restriction levels. The following is a sample sequence of operations illustrating how our middleware works to enforce the security property.

\textbf{Stage 1:} The first considered user-initiated request is creating a VM, namely \texttt{leaps_vm}.

\textbf{Stage 2:} User requests for attaching the VM to security group \textit{no_connection}. Security group \textit{no_connection} hypothetically implements the most restricted security policy. The request passes through Nova pipeline and is sent to OpenStack networking service (Neutron) to be executed. Figure 4 shows the corresponding logged request made to Neutron and logged port update event for attaching security group, which means the request has reached Neutron service.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image1}
\caption{Neutron service logged actions following receiving the request from Nova}
\end{figure}

Figure 5 shows logged response to the underlined API request, which is associated with attaching to security group \textit{no_connection}. The status code 202 shows the request has been accepted and successfully processed.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image2}
\caption{Logged response with status code 202, associated with request for attaching to security group \textit{no_connection}}
\end{figure}

\textbf{Stage 3:} The cloud user starts the VM. \textit{PERMON} intercepts the request while it is passing through the pipeline and inspects its body. The request body is the content of \texttt{wsgi.input} of \texttt{environ} argument that is sent to the server. If \textit{PERMON} finds the body is associated with starting a VM through verifying \texttt{wsgi.input} attribute of a passing request, it extracts the ID of the VM from another attribute of that request, \texttt{PATH_INFO}. Next, with the extracted ID, \textit{LeaPS} queries the Neutron database for the security groups \texttt{leaps_vm} is attached to. According to these currently attached security groups, it populates the watchlist with the allowed security groups for the started VM.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image3}
\caption{Logged response with status code 202, associated with request for attaching to security group \textit{no_connection}}
\end{figure}

Figure 6 illustrates the content of the intercepted parameter, \texttt{wsgi.input}. The format of \texttt{PATH_INFO} and the content of the request corresponding to different server actions are interpreted according to the API documentation as follows. \texttt{PATH_INFO} of a request for starting a VM is in the format of \texttt{/servers/server_id/action}. The format provided in the API documentation is used to identify requested actions and to extract different parameters e.g., \texttt{server_id}. As it is shown in Figure 7, we keep track of the running VM with an index to the security groups it can be legitimately attached to.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image4}
\caption{Intercepted request body for starting VM}
\end{figure}
Stage 4: At this point, the attacker tries to downgrade the security group of the running VM, \textit{leaps\_vm}, by attaching it to a less restricted security group, namely \textit{essential} (figure 8).

The content of the request is inspected by \textit{PERMON}. Having found the body corresponding to a security group attachment, \textit{PERMON} extracts the security group name and the ID of the VM, which LeaPS looks for and finds among IDs of running VMs in its database. This means that \textit{leaps\_vm} can only be attached to more restricted security groups than what it is currently attached to, i.e., \textit{no\_connection}. LeaPS goes through the legitimate security groups indexed by \textit{leaps\_vm}. As security group \textit{essential} is not among them, LeaPS decides on denial of the request.

Consequently, \textit{PERMON} refuses to pass this request to the next element in the pipeline, and prepares its own response with status code 403, associated with request for attaching to security group \textit{essential} made to Nova, which shows the request has been blocked by our middleware.

VI. DISCUSSIONS

Benefit to Log Analysis In addition to security policy enforcement, the \textit{PERMON} middleware can also benefit log processing. In general, OpenStack services only log their responses to the received requests. For log analyses which aim at identifying the triggering requests, those logged responses may not provide sufficient information. Specifically, the log entries can be parsed to extract different fields among which \textit{METHOD} and \textit{PATH\_INFO} are supposed to map to different types of requested events. However, as mentioned earlier, many events cannot be distinguished from each other. For example, all requests to invoke a server to take some specific action will have the same \textit{METHOD} and \textit{PATH\_INFO}, which renders log interpretation infeasible. Table II shows three examples of logged responses to such requests. Except for the VM ID at the end of \textit{PATH\_INFO}, which is not useful for identifying the type of requested event, all three log entries have the same combination of \textit{METHOD} and \textit{PATH\_INFO}. To this end, our \textit{PERMON} middleware logs and examines the intercepted body of requests in order to identify the type of logged requests. This provides a feasible solution for more accurate log processing.

**TABLE II**

<table>
<thead>
<tr>
<th>OpenStack Log Entry</th>
<th>Event Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>POST /v2.1/a6627ffa0c4f4a3ebaefe05c0b93f4c6/servers/f6128951-0c48-4a11-bb8b-5e96da77b698/</td>
<td>Start VM</td>
</tr>
<tr>
<td>POST /v2.1/a6627ffa0c4f4a3ebaefe05c0b93f4c6/servers/1223d052-bc35-485a-9b31-1830bca80fd7/</td>
<td>Stop VM</td>
</tr>
<tr>
<td>POST /v2.1/a6627ffa0c4f4a3ebaefe05c0b93f4c6/servers/4c886192-43ad-4598-90dd-34e24c84fcd9/</td>
<td>Add Security Group</td>
</tr>
</tbody>
</table>

Discrepancy between Console and GUI Requests By studying intercepted attributes, we note that wsgi\_input and \textit{METHOD} of an issued request can be different depending on whether the request is made from command-line or inside the OpenStack dashboard. For instance, Table III shows the content of two intercepted attributes logged by \textit{PERMON} when a request for attaching VM1 to security group SG1 is made from command-line as opposed to when it is issued from OpenStack dashboard. The logged attributes are compatible.
with what is indicated in the API documentation for attaching a VM to a security group in the former case; however, they are mapped to API calls for updating a VM indicated in the API documentation in the latter case. In this work, we set our matching rules according to the API documentation and we expect such discrepancy between internally generated parameters corresponding to identical requests to be addressed in future OpenStack releases.

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>DISCREPANCY BETWEEN REQUESTS MADE FROM OPENSTACK COMMAND-LINE AND DASHBOARD</th>
</tr>
</thead>
</table>
| Dashboard | PUT  
| wsgi.input | [
| | server: [
| | name: 
| | VM1]] | Update VM |
| Command-line | POST  
| | addSecurityGroup  
| | [
| | name: 
| | "SG1"] | Add security group |

VII. Conclusion

This paper presented a security policy enforcement middleware, PERMON, which was designed as a pluggable module in OpenStack Nova service. PERMON provided control over user-initiated requests according to given security policies or properties. Working along with our proactive verification module LeaPS, PERMON could make decisions about either allowing or denying a request in an efficient manner with only negligible delay to legitimate users. Furthermore, by inspecting the request body to identify requests that are otherwise not distinguishable, PERMON could also bring added value to log analysis in OpenStack. Our future work will focus on evaluating the performance of PERMON through experiments. Moreover, current implementation of PERMON is OpenStack specific, but we explore the possibility of integrating it to other cloud platforms.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their valuable comments, as well as Yosr Jarraya and Makan Pourzandi for their contributions to the paper. This work is partially supported by the Natural Sciences and Engineering Research Council of Canada and Ericsson Canada under CRD Grant N01566, Discovery Grant N01035 and Prompt Quebec.

REFERENCES


