UDC 004.056

Aksyonenko Ilya,
Pavlo Rehida

APPLICATIONS OF SEQUENCE-TO-SEQUENCE AUTOENCODER NETWORKS IN REQUEST ANOMALY DETECTION

Аксьоненко Ілля,
Павло Регіда

ЗАСТОСУВАННЯ SEQUENCE-TO-SEQUENCE AUTOENCODER НЕЙРОННИХ МЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ АНОМАЛЬНИХ ЗАПИТІВ

This paper provides an insight into utilizing machine learning techniques to improve web application firewall (WAF) performance. A brief overview of existing techniques is provided, and a solution is proposed to optimize security breach alerts and anomaly detection capabilities of WAF software. An existing seq2seq autoencoder architecture is applied to solve the problem of efficient attack detection in WAF software.

Key words: WAF, LSTM, seq2seq, autoencoder.
Fig.: 4. Tabl. 0. Bibl. 0.

У статті розглядається використання технологій машинного навчання для підвищення ефективності web application firewall (WAF). На основі існуючої архітектури та рішень пропонується новий метод розпізнавання аномалій та атак. Нейронна мережа з архітектурою seq2seq використовується для вирішення задачі ефективного розпізнавання атак у послідовностях символів.

Ключові слова: WAF, LSTM, seq2seq, autoencoder.
Рис.: 4. Табл. 0. Бібл. 0.

Relevance analysis. WAF usage is becoming a staple in securing web applications, being required by industry-leading data protection standards, such as PCI DSS. Most of the provided solutions, however, rely on blacklisting [1,2] malicious requests either via regular expressions or attack fingerprints.

Although effective to some extent, these protection methods can be bypassed, which has happened in the past, as in the example [3]. Moreover, they by design can only prevent against known attacks, as they do not have measures past basic heuristics to prevent attacks previously unknown to them. Thus, a more general approach based on generic anomaly detection was proposed [4], and our method is an extension of this idea.

WAF general architecture. According to OWASP, a web application firewall (WAF) is an application firewall for HTTP applications. It applies a set of rules to an HTTP conversation. [5] Generally, these rules cover common attacks such as cross-site scripting (XSS) and SQL injection. As a WAF can be considered a filtering reverse proxy, its simplified and generalized data flow diagram could be visualized as follows:

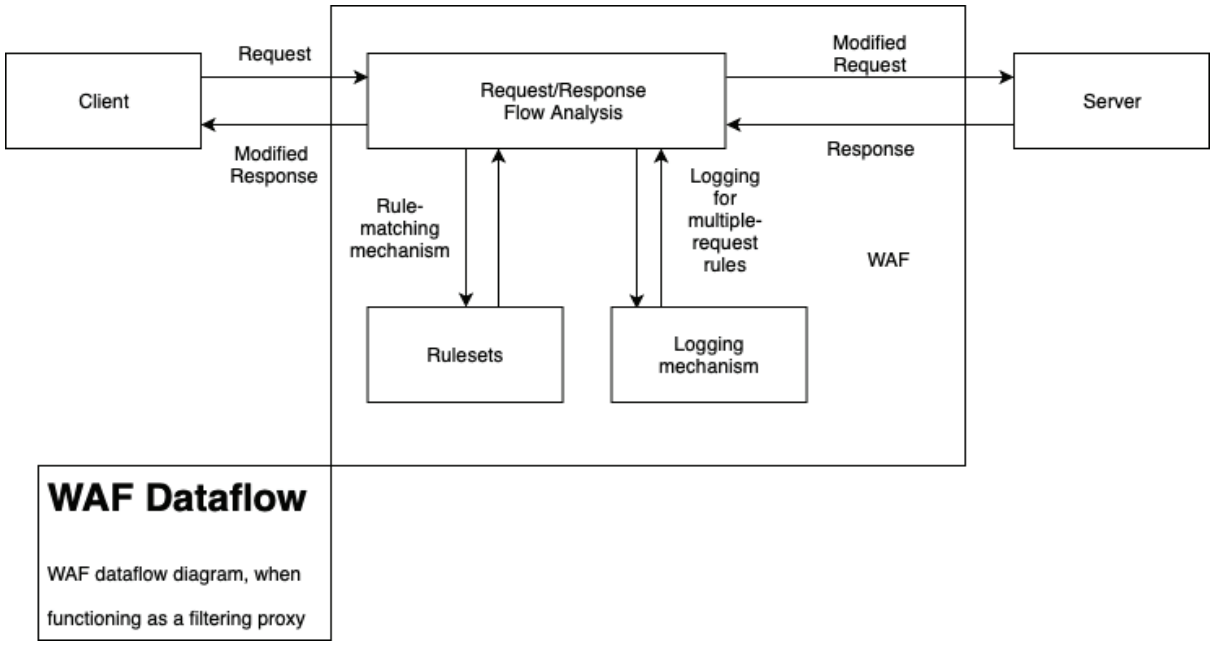

Fig.1. Generic WAF dataflow diagram

The proposed solution does not change this basic data flow, however, regular-expression based rulesets are replaced with a seq2seq autoencoder neural network that learns on previous trusted user requests. This also allows to implement our solution on top of existing software, combining machine learning-based anomaly detection with a vast range of existing attack fingerprints and rulesets.

Overview of existing solutions. This approach is mostly based on the work described above [3], as it pioneered the method. The idea is to implement a sequence to sequence autoencoder neural network with the following configuration:

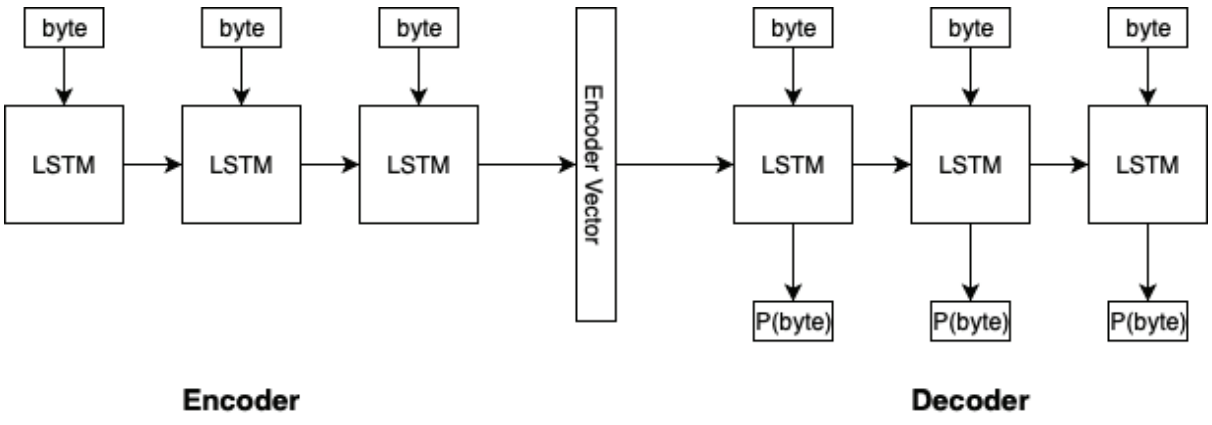

Fig. 2. Simplified LSTM network architecture

Figure 2 illustrating how the sequence is processed by the encoder and how the encoder internal state is used to initialize the decoder. The decoder output is used to determine estimated probabilities for sequence symbols, with anomalous parts of the sequence having significantly lower probabilities.

The network consists of the encoder and decoder LSTM networks [6], both trained with the exact same dataset of legitimate requests to the application. The internal state of the encoder, a vector of fixed length, is then used to initialize the decoder. The decoder is then used to determine the probability of the next symbol in the sequence, as it is trained to reconstruct known sequences from the input vector. Thus, if the request is anomalous, the probabilities of symbols in the anomalous (previously unseen) part of the request are significantly lower than average, allowing to highlight possible payloads for vulnerability exploitation. [4]

As described in the article above, the detection process can be visualized by the following example:

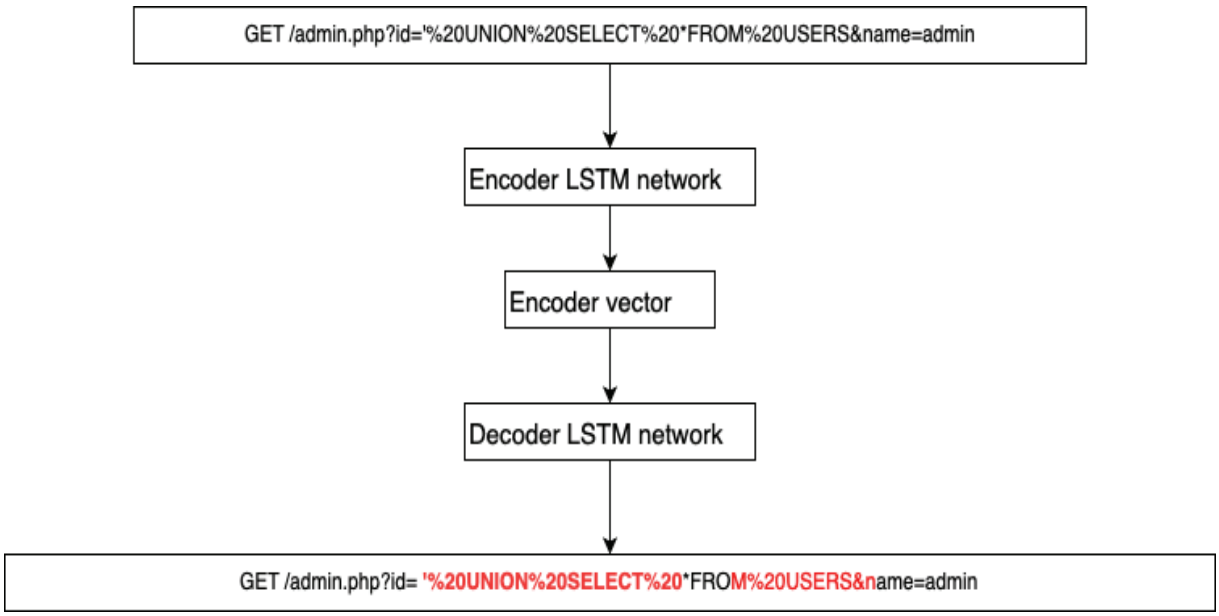

Fig.3. An implemented method to detect HTTP request anomalies.

The example details that the process does not give exact results, only estimating probability thresholds for an anomalous sequence.

Our approach. However, the approach described in the article can be optimized. The algorithm suggested by Alexandra Murzina, Irina Stepanyuk, Fedor Sakharov, and Arseny Reutov trains both networks on the whole request string, and thus, more time is required to allow the network to recognise HTTP request patterns.

Instead, we propose field-based anomaly detection, which adds an extra step to the detection flow. Instead of training one anomaly detection seq2seq autoencoder network, we propose to add a layer that separates HTTP form field values and supplies them both as training data and input for the several detectors, identical in structure. Field variable names can be also whitelisted, as the list of legitimate request fields is

known before the implementation of the WAF. With this approach, it is possible to achieve faster and more specific training, as each network will learn to only reproduce values of a single field.

With our solution, the example above is solved in another way:

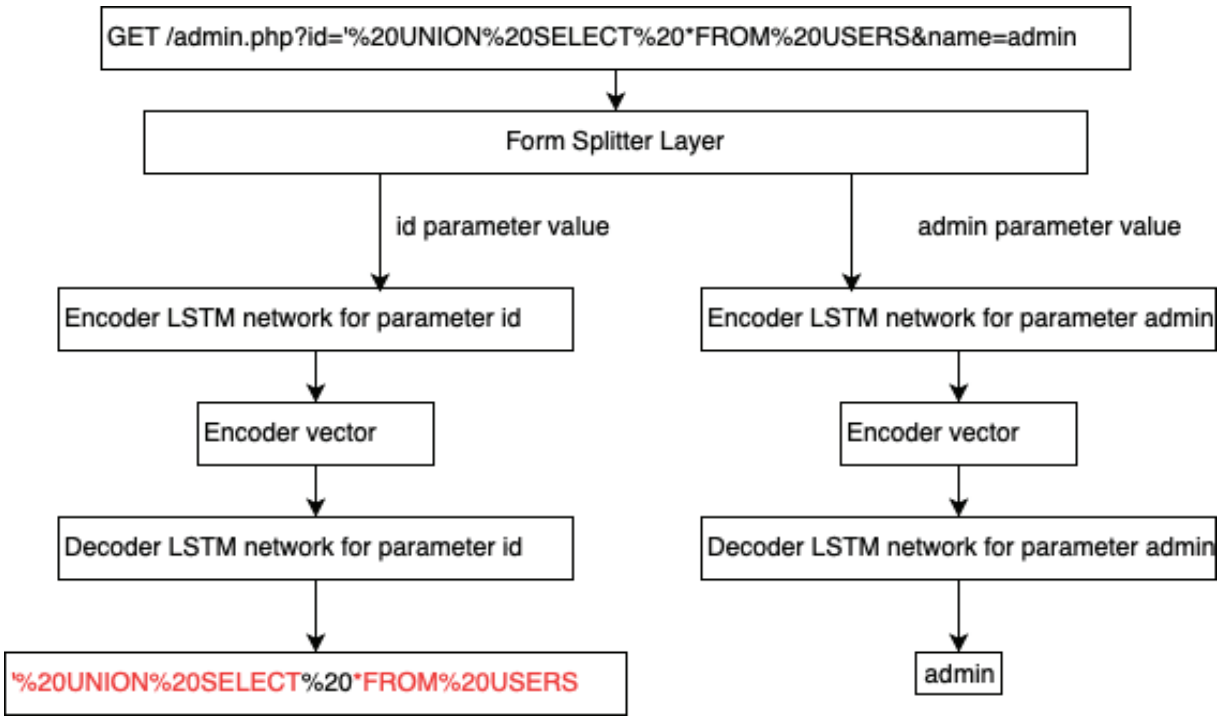

Fig. 3. An example of request processing with our approach

Although the method is still the same, the request-splitting approach allows to pinpoint the form field where the injection takes place.

In general, the algorithm proposed gives up multiple-parameter anomaly detection in favor of faster training and better flexibility (as when parameter submission form is changed, the trained networks can be reused to validate form values).

It could be best applied together with conventional attack-detection methods. As our method allows for localized detection, as opposed to highlighting anomalous parts of the request string, it is possible to use it as a trigger for a conventional WAF mechanism to use an extended set of heuristics to check an anomalous request.

Also, the approach could be used as a standalone solution, alerting the security personnel and automatically blocking users that send a high number of potentially anomalous requests in each timeframe.

Additionally, it can be used only as a monitoring solution, inspecting the mirrored traffic to/from an application and raising alerts based on detected anomalies.

Conclusions. In general, the idea proposed is a tradeoff between multi-value anomalies and faster, more specific detection. This theoretical approach can be the basis to implement practical machine learning in the field of web application firewall applications.

References

1. Nginx WAF. https://docs.nginx.com/nginx-waf. Accessed 12 May 2019
2. Mod_security documentation. http://modsecurity.org/rules.html. Accessed 12 May 2019
3. Web Application Firewall (WAF) Evasion Techniques #2. https://medium.com/secjuice/web-application-firewall-waf-evasion-techniques-2-125995f3e7b0. Accessed 12 May 2019
4. A. Murzina, I. Stepanyuk, F. Sakharov, A. Reutov. Detecting Web Attacks with a Seq2Seq Autoencoder. https://habr.com/en/company/pt/blog/441030/. Accessed 12 May 2019
5. Open Web Application Security Project documentation. https://www.owasp.org/index.php /Web_Application_Firewall. Accessed 12 May 2019
6. A Gentle Introduction to LSTM Autoencoders. https://machinelearningmastery.com/lstm-autoencoders/. Accessed 12 May 2019

Довідка про авторів

Аксьоненко Ілля Олегович – студент, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

E-mail: ilya.aksyonenko@gmail.com

Aksyonenko Ilya – student, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

Регіда Павло Генадійович – асистент, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

E-mail: pavel.regida@gmail.com

Rehida Pavlo – Assistant Professor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

РОЗШИРЕНА АНОТАЦІЯ

Аксьоненко Ілля,
Павло Регіда

ЗАСТОСУВАННЯ SEQUENCE-TO-SEQUENCE AUTOENCODER НЕЙРОННИХ МЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ АНОМАЛЬНИХ ЗАПИТІВ

Актуальність теми дослідження. Проблема розпізнавання веб-атак стає більш актуальною в останні дні у зв'язку зі зростаючою долею застосунків, що використовують веб-технології. Таким чином, проблемою є створення універсального способу розпізнавання атак, яка не буде базуватися на фільтрах. Дана робота присвячена проблемі розпізнавання веб-атак як аномалій за допомогою нейронних мереж.

Постановка проблеми. Неефективність та недоліки існуючих систем розпізнавання веб-атак,

Аналіз останніх досліджень і публікацій. Робота побудована на і є поширенням ідеї використаня seq2seq мереж для розпізнавання аномалій. Ідея була представлена у статті A. Murzina, I. Stepanyuk, F. Sakharov, A. Reutov: *Detecting Web Attacks with a Seq2Seq Autoencoder.*

Виділення недосліджених частин загальної проблеми. Дана стаття присвячена розробці додаткового шару обробки даних до нейронної мережі для оптимізації певних характеристик системи.

Постановка завдання. Завданням є створити теоретичну архітектуру, що базується на існуючих моделях, але використовує підхід поділу вхідних даних до обробки них мережею.

Викладення основного матеріалу. Проведено аналіз seq2seq підходу до розпізнавання веб-атак, запропоновано новий метод розділення даних по параметрам запиту.

Висновки. За результатами аналізу було зроблено висновок, що запропонований підхід має переваги для певних сценаріїв застосування і може використовувватися разом з існуючими методами захисту.

Ключові слова: WAF, LSTM, seq2seq, autoencoder.