Anomaly Detection System Using Beta Mixture Models and Outlier Detection

Nour Moustafa, Gideon Creech and Jill Slay

Abstract An intrusion detection system (IDS) plays a significant role in recognising suspicious activities in hosts or networks, even though this system still has the challenge of producing high false positive rates with the degradation of its performance. This paper suggests a new beta mixture technique (BMM-ADS) using the principle of anomaly detection. This establishes a profile from the normal data and considers any deviation from this profile as an anomaly. The experimental outcomes show that the BMM-ADS technique provides a higher detection rate and lower false rate than three recent techniques on the UNSW-NB15 data set.

Keywords Intrusion detection system (IDS) · Anomaly detection system (ADS) Beta mixture model (BMM) · Outlier detection

1 Introduction

An intrusion detection system (IDS) has become an essential application to defend against cyber attackers. The methodologies of IDS can be categorised into misuse-based, anomaly-based or hybrid of the previous two [2, 16]. On the one hand, a misuse-based IDS monitors the activities of hosts or networks to match observed instances with a well-known blacklist in which includes the existing signatures of known attacks. Though this method provides higher detection rates (DR) and lower false positive rates (FPR), it cannot detect new attacks (i.e. zero-day attacks). Additionally, it demands a huge effort to regularly update its blacklist with the new rules of

N. Moustafa (✉) · G. Creech · J. Slay
The Australian Centre for Cyber Security, University of New South Wales, Canberra, Australia
e-mail: nour.moustafa@unsw.edu.au

G. Creech
e-mail: G.Creech@adfa.edu.au

J. Slay
e-mail: j.slay@adfa.edu.au

P. K. Pattnaik et al. (eds.), *Progress in Computing, Analytics and Networking*, Advances in Intelligent Systems and Computing 710,
https://doi.org/10.1007/978-981-10-7871-2_13

suspicious activities [3]. An anomaly-based IDS, on the other hand, constructs a profile from legitimate data and detects any variation from the profile as an anomaly. This method can identify existing and zero-day attacks, so it will be better than misuse–based if its potential procedures are successfully designed. However, constructing a normal profile is very difficult due to the difficulty of involving all possible patterns of normal data [11].

Therefore, we propose constructing a normal profile using statistical models, in particular a beta mixture model (BMM) for several reasons [4, 8, 15]. Firstly, statistical models can simply determine potential properties of network patterns for both features and vectors [14]. Secondly, mixture models can precisely fit Gaussian and non-Gaussian data with specifying data edges. This means that any data outside of these edges will be handled as outliers/anomalies. Thirdly, a BMM can be designed by scaling data edges between a finite range ($[x, y], x, y \in R$) in order to control data boundaries within this range.

In this paper, we suggest an anomaly-based IDS based on the theory of beta mixture models in order to establish a profile from normal data. To recognise suspicions observations, we propose a decision-making method for detecting existing and new attacks using a baseline of the lower-upper interquartile range (IQR) [17]. This method measures the lower and upper boundaries of the normal profile and treats any observation outside of this range as an anomaly. The proposed BMM-ADS technique is evaluated on the UNSW-NB15 data set[1] [13], providing a higher DR and lower FPR than three compelling techniques.

The rest of this paper is organised as follows. Section 2 explains the background and related studies to the IDS technology. The new anomaly detection system based on the beta mixture model is explained in Sect. 3. The experimental results and discussions are provided in Sect. 4. Finally, we summarise the paper.

2 Background and Previous Work

An intrusion detection system (IDS) is a mechanism for monitoring host or network activities to recognise possible threats by estimating their vulnerabilities of Confidentiality, Integrity and Availability (CIA) principles [12, 15, 16, 20]. There are two kinds of IDSs depending on the data source: a host-based IDS inspects the activities of a computer system by accumulating information which take place in a client system, whereas a network-based IDS monitors network traffic to define network attacks that happen throughout that network [15].

An anomaly-based IDS (ADS) is a type of IDS for monitoring events that happen in a host or network to recognise possible threats. A classic ADS comprises three components of a data source, data pre-processing and decision-making method. The data source involves data gathered from host traces or network traffic, while the data

[1] "The UNSW-NB15 data set", https://www.unsw.adfa.edu.au/australian-centre-for-cyber-security/cybersecurity/ADFA-NB15-Datasets/, January 2017.

pre-processing includes the construction of attributes from the data that are then sent to the detection method, which is utilised for identifying malicious activities [3, 11]. This technique establishes a profile from legitimate patterns and considers the variations from this profile as attacks [15]. Nevertheless, identifying the boundaries of such a profile and the method of recognising outliers is still challenging [2, 4, 11, 14].

Several studies have been conducted to address this challenge. For example, Greggio [5] developed ADS based on the Gaussian Mixture Model. The mixture component was specified by estimating the parameters of the normal data and handling any data outside of this range as anomalies. Tan et al. [20] suggested a multivariate correlation technique for establishing a DoS identification mechanism using a triangular method of the lower correlation matrix used for estimating the correlation between attributes in order to help in identifying malicious instances.

Fan et al. [7] proposed a Bayesian inference method for designing a network collaboration framework for data via gathering feedback from distributed nodes and modelled by a beta distribution to classify the error rates for different ADS techniques. Singh et al. [19] suggested a distributed ADS based on the random forest algorithm for identifying botnets from a large-scale network. Fortunati et al. [6] proposed a statistical ADS using a generalised version of the inequality for random observation. The results of this technique slightly improved the accuracy detection using the KDD99 data set.

Most of the studies above enhanced the detection rate because they used a particular baseline/threshold in the classification stage that would be either a binary value, which is 1 for attack and 0 for normal, or static value that did not estimate from real network environments. Nevertheless, the results were often biased towards normal observations that provided high FPRs [9]. In our recent work [15], we developed a Geometric Area Analysis mechanism using trapezoidal area estimation for each instance calculated from the BMM parameters for network attributes and the distances between instances, but this paper proposes estimating the baseline from the processed network data with flexible inference overlays using a BMM-ADS in order to improve the DR and decrease the FPR.

3 Beta Mixture Model-ADS

The mixture technique is a robust probabilistic model for representing a subset multivariate data that demonstrates the whole data set. The beta mixture model (BMM) precisely fits the bounded property data with less complexity than the Gaussian mixture model (GMM) [10]. However, the GMM can model any random distribution with appropriate mixture components. There are some components do not correctly characterise boundaries when testing data are bounded or semi-bounded [4].

The features of network data cannot accurately fit a normal distribution because they do not fit its unbounded and symmetric edges (i.e.] $-\infty, \infty$[) [14]. As in the

data sets of NSLKDD[2] and UNSW-NB15, their features can be represented in a semi-bounded range of $[0, N]$, such that N denotes an asymmetric number. A beta distribution can fit data in a more elastic form than a normal distribution and models arbitrary features that have a finite range of ($[x, y], x, y \in R$), such as $[0, 1]$. Consequently, we use BMM for building the normal profile of ADS [8, 10].

A beta distribution's probability density function (PDF) is calculated by

$$Beta(x; \upsilon, \omega) = \frac{1}{beta(\upsilon, \omega)} x^{\upsilon-\omega}(1-x)^{\omega-1}, \upsilon, \omega > 0 \tag{1}$$

such that x is the random variables/attributes, $beta(\upsilon, \omega)$ is the beta function, $beta(\upsilon, \omega) = \Gamma(\upsilon)\Gamma(\omega)/\Gamma(\upsilon + \omega)$, υ and ω refer to the shaped parameters that model the beta distribution, and $\Gamma(.)$ denotes the gamma function $\Gamma(c) = \int_0^\infty \exp(-t)t^{c-1}dt$.

In our new BMM-ADS technique, a BMM is used for estimating the network feature's PDFs. It is noted that network features are independent [20], while multivariate attributes are dependent in many situations. Nonetheless, for any attribute (x) containing L values, the dependence between values $x_1, \ldots, x_L$ is indicated using a mixture technique even if each component can design observations with independent attributes. We declare the PDF multivariate BMM for some observations as

$$\begin{aligned} f(x; \pi, \nu, \omega) &= \sum_{i=1}^{I} \Pi_i Beta(X, \nu_i, \omega_i) \\ &= \sum_{i=1}^{I} \Pi_i \prod_{l=1}^{L} Beta(x_l, \nu_{li}, \omega_{li}) \end{aligned} \tag{2}$$

where I indicates component number ($X = \{x_1, ..., x_L\}$, $\prod = \{\Pi_1, .., \Pi_I\}$, $\upsilon = \{\upsilon_1, ..., \upsilon_I\}$, $\omega = \{\omega_1, ..., \omega_I\}$), Π_i refers to the mixing component (where $\sum_{i=1}^{I} \Pi_i = 1, 0 < \pi < 1$), $\{\upsilon_i, \omega_i\}$ are the parameter instances of the i^{th} mixture component, $Beta(X; \upsilon_i, \omega_i)$ is the component parameters, and $\{\upsilon_{lI}, \ldots, \omega_{lI}\}$ indicate the beta parameters for attribute x_l.

To explain the BMM, given two random variables (x_1 and x_2), their parameters are computed using the EM technique, as detailed in [10]. Figure 1 shows an example for modelling two variables by BMM, where parameters of $x_1(\pi, \nu, \omega)$ be $(0.55, 30, 10)$ and parameters of x_2 be $(0.45, 10, 30)$. We estimate the BMM parameters for the data set features in order to construct a normal profile, which has a wide range of PDFs that could represent the entire observations of normal behaviours.

The learning process of BMM is a significant task for estimating the parameters and selecting the number of components (M). We use the maximum likelihood suggested in [10] to estimate the parameters of the finite BMM and choose the number of components.

[2]"NSLKDD data set", https://web.archive.org/web/20150205070216/, http://nsl.cs.unb.ca/NSL-KDD/, January 2017.

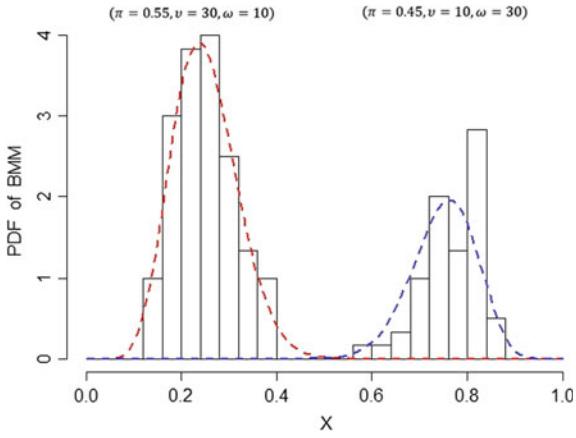

Fig. 1 BMM for two random variables

In this study, we suggest a new BMM-ADS technique for recognising anomaly instances. In the training phase of the technique, we establish the legitimate profile using BMM parameters, PDFs and a lower-upper IQR baseline for learning legitimate network data, whereas the abnormal instances which are outside of the baseline are considered as suspicious instances in the testing phase, as detailed in the following two sections.

3.1 Training phase

The BMM-ADS technique has to learn using purely legitimate observations in order to make sure that the technique can correctly detect malicious ones. Given a set of normal observations ($r_{1:n}^{normal}$) in which each vector consists of a set of features, where $r_{1:n}^{normal} = \{x_1, x_2, \ldots, x_D\}^{normal}$, the legitimate profile involves only statistical measures from $r_{1:n}^{normal}$. They involve the estimated parameters (π, v, ω) of the BMM to calculate the PDF of the beta distribution ($Beta(x; \pi, v, \omega)$) for each vector in the training set.

Algorithm 1 presents the suggested process for establishing a legitimate profile (pro) using the parameters of the BMM estimated for all the legitimate instances $r_{1:n}^{normal}$ using the equations proposed in [10], and then the PDFs of the features ($x_{1:D}$) are computed using Eq. 2. After that, IQR is calculated by subtracting the first quartile from the third quartile of the PDFs to specify a baseline for identifying suspicious observations in the testing phase. Quartiles can divide a range of data into contiguous intervals with equal probabilities [17].

Algorithm 1 Normal profile construction of normal instances

Input: normal observations ($r_{1:n}^{normal}$)
Output: normal profile (pro)

1: **for** each record i in ($r_{1:n}^{normal}$) **do**
2: calculate the parameters ($\pi_i, v_{,i}, \omega_i$) of the BMM as in [14]
3: calculate the PDFs using equation 2 using the parameters of Step 2
4: **end for**
5: calculate $lower = quartile(PDFs, 1)$
6: calculate $upper = quartile(PDFs, 3)$
7: calculate $IQR = upper - lower$
8: pro $\leftarrow \{(\pi_i, v_{,i}, \omega_i), (lower, upper, IQR)\}$
* **return** pro

3.2 Testing Phase and Attack Detection

For testing each observed record, the Beta PDF ($PDF^{testing}$) of each instance ($r^{testing}$) is calculated using the same parameters of the normal profile (pro). Algorithm 2 describes the steps in the testing phase and decision-making method for recognising the Beta PDFs of the malicious records, with step 1 describing the PDF of each observed instance using the normal parameters ($\pi_i, v_{,i}, \omega_i$).

Algorithm 2 Testing phase and decision-making method

input : observed record ($r^{testing}$), pro
output : normal or attack

1: calculate the $PDF^{testing}$ using equation 2 using the parameters ($\pi_i, v_{,i}, \omega_i$)
2: **if** ($PDF^{testing} < (lower – w * (IQR))$ || ($PDF^{testing} > (upper + w * (IQR))$ **then**
3: **return** attack
4: **else**
5: **return** normal
6: **end if**

Steps 2 to 6 explain the steps of the decision-making method. The IQR is the length of the box in the box-and-whisker plot, specifying outliers as values that locate more than 1.5 the length of the box from either end of the box. In more detail, the IQR of the normal instances is calculated for identifying the anomalies of any observed record ($r^{testing}$) in the testing phase which is treated as any instance located below ($lower – w * (IQR)$) or above ($upper + w * (IQR)$), such that w refers to the interval values between 1.5 and 3 [17]. The decision of detection depends on considering any $PDF^{testing}$ falling outside of this range as a malicious record, otherwise normal.

4 Empirical Results and Discussion

4.1 Evaluation Criteria

Multiple experiments were conduced on the UNSW-NB15 data set in order to appraise the performance of the BMM-ADS technique using the metrics of accuracy, DR, FPR and ROC curves, defined as in the following points.

- The **accuracy** is the proportion of all legitimate and malicious observations correctly categorised, that is,

$$accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{3}$$

- The **detection rate (DR)** is the proportion of correctly identified malicious observations, that is,

$$DR = \frac{TP}{(TP + FN)} \tag{4}$$

- The **false positive rate (FPR)** is the proportion of incorrectly identified malicious observations, that is,

$$FPR = \frac{FP}{(FP + TN)} \tag{5}$$

where TP (true positive) refers to the number of actual malicious observations categorised as attacks, TN (true negative) indicates the number of actual normal records categorised as normal, FP (false positive) means the number of actual normal records categorised as attacks, and FN (false negative) refers to the number of actual malicious observations categorised as normal.

4.2 Description of Pre-processing Stage

The UNSW-NB15 data set was used for evaluating the effectiveness of the proposed BMM-ADS technique, which has a collection of recent normal and attack observations. Its size is nearly 100 GBs extracted 2,540,044 records, which are kept in four CSV files. Each record includes 47 attributes and its label. It includes ten different classes, one legitimate and nine kinds of malicious events. A part of the data set is prepared for training and testing NIDS techniques in [14]. The proposed technique was assessed using eight features selected from the UNSW-NB15 using the principal component analysis technique listed in Table 1.

Table 1 Feature selected from UNSW-NB15 data set

Data set	Selected features
UNSW-NB15	ct_dst_sport_ltm, tcprtt, dwin, ct_src_dport_ltm, ct_dst_src_ltm, ct_dst_ltm, smean, service

In order to carry out the experiments, arbitrary samples are selected from the UNSW-NB15 data set with sizes vary between 50,000 and 200,000. In each one, legitimate instances were approximately 55–65% of the total size, with some used to create the legitimate profile and the testing set.

5 Empirical Results

The performance of the BMM-ADS mechanism was evaluated using the overall accuracy, DR and FPR on the feature adopted from the UNSW-NB15 data set, demonstrated in Table 2. Furthermore, the ROC curves which represent the relationship between the DRs and FPRs with different w values are presented in (Fig. 2). The DR and accuracy increased from 82.4% to 92.7% and 84.2% and 93.4%, respectively; however, the FPR decreased from 10.3% to 5.9 % while the w value increased from 1.5 to 3.

Table 3 shows that the proposed mechanism identified observation types of the UNSW-NB15 data set with normal DRs fluctuating between 83.4% and 94.0% when the w value increased from 1.5 to 3. Likewise, the DRs of the malicious kinds increased gradually from an average of 35.7% to an average of 89.6%.

Some attack types achieved higher DRs within the gradual increase of the w value, while others do not produce high DRs due to the small similarities between malicious and legitimate observations. Since the UNSW-NB15 data set is similar to real networks with broad variations of legitimate and malicious patterns, applying a feature reduction method could make a clear difference between these patterns, improving the performance of the proposed technique. We observe that the variances of the selected feature are close, leading an overlap the PDFs of the attacks in normal ones.

Table 2 Performance of features selected from UNSW-NB15 data set

w value	DR (%)	Accuracy (%)	FPR (%)
1.5	82.4	84.2	10.3
2	84.5	86.3	8.8
2.5	90.5	91.5	7.2
3	92.7	93.4	5.9

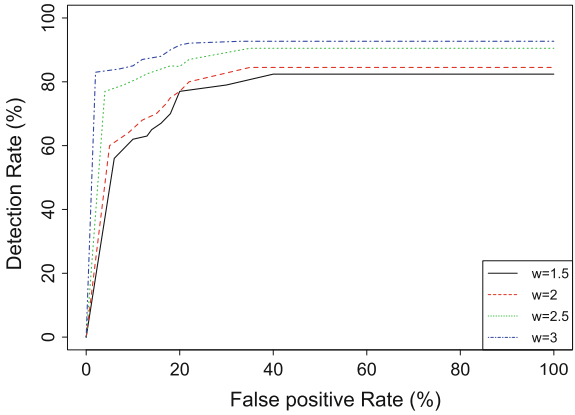

Fig. 2 ROC curves of UNSW-NB15 data set with w values

Table 3 Comparison of DRs (%) on UNSW-NB15 data set

	w values			
Record type	1.5 (%)	2 (%)	2.5 (%)	3 (%)
Normal	81.2	85.4	90.5	93.4
DoS	82.6	85.3	86.1	89.6
Backdoor	55.3	61.2	62.3	63.8
Exploits	60.2	67.1	73.6	79.4
Analysis	72.6	71.2	77.1	83.4
Generic	80.5	86.3	86.3	86.3
Fuzzers	42.4	50.1	50.8	52.8
Shellcode	42.2	44.3	47.2	48.7
Reconnaissance	50.8	54.2	54.2	55.6
Worms	35.7	40.3	42.2	47.8

We compare our proposed technique with three recent techniques, namely Multivariate Correlation Analysis (MCA) [20], Artificial Immune System (AIS) [18] and Filter-based Support Vector Machine (FSVM) [1] on the UNSW-NB15 data set. As listed in Table 4, the findings obviously show the superiority of our mechanism in terms of detection and false positive rates. This is because our technique is designed to model the normal data with a flexible shape, which includes a wide range of normal PDFs, and the decision method of IQR can therefore find the outliers from the profile as anomalies.

The MCA technique depends on only finding correlations between features with the Gaussian mixture model to recognise the DoS attacks, which sometimes cannot specify accurate edges between normal and attack PDFs. The other two techniques rely on learning normal and abnormal data in the training stage, which is the

Table 4 Comparison of performance of four techniques

Technique	DR (%)	FPR (%)
MCA [20]	88.3	11.6
AIS [18]	83.5	15.7
FSVM [1]	90.4	8.5
BMM-ADS	92.7	5.9

principle of rule-based. Such techniques demand a huge number of instances to be properly learned which makes it in online learning. Although these techniques reflected a higher performance evaluation on the outdated KDD99 data set or its improved version NSLKDD, our technique outperforms them in terms of DRs and FPRs. This is an indication that our technique can achieve better than these mechanisms on real network data, as it is hard to receive all security events and malware at the same time from different nodes.

6 Conclusion

This paper covers a proposed anomaly detection system based on the beta mixture model for establishing a profile from normal network data. In order to recognise malicious observations, we suggest the lower-upper interquartile threshold as a baseline of legitimate profile and any variations from this threshold are considered as an attack. The experimental results showed the higher performance evaluation of this technique and its superiority compared with three recent mechanisms. In future, we are planning to investigate feature reduction methods to find clear differences between selected features, further improving the performance of these techniques.

References

1. Ambusaidi, M.A., He, X., Nanda, P., Tan, Z.: Building an intrusion detection system using a filter-based feature selection algorithm. IEEE transactions on computers **65**(10), 2986–2998 (2016)
2. Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.K.: Network anomaly detection: methods, systems and tools. IEEE communications surveys & tutorials **16**(1), 303–336 (2014)
3. Creech, G., Hu, J.: A semantic approach to host-based intrusion detection systems using contiguousand discontiguous system call patterns. IEEE Transactions on Computers **63**(4), 807–819 (2014)
4. Escobar, M.D., West, M.: Bayesian density estimation and inference using mixtures. Journal of the american statistical association **90**(430), 577–588 (1995)
5. Fan, W., Bouguila, N., Ziou, D.: Unsupervised anomaly intrusion detection via localized bayesian feature selection. In: Data Mining (ICDM), 2011 IEEE 11th International Conference on, pp. 1032–1037. IEEE (2011)

6. Fortunati, S., Gini, F., Greco, M.S., Farina, A., Graziano, A., Giompapa, S.: An improvement of the state-of-the-art covariance-based methods for statistical anomaly detection algorithms. Signal, Image and Video Processing **10**(4), 687–694 (2016)
7. Fung, C.J., Zhu, Q., Boutaba, R., Ba, T., et al.: Bayesian decision aggregation in collaborative intrusion detection networks. In: Network Operations and Management Symposium (NOMS), 2010 IEEE, pp. 349–356. IEEE (2010)
8. Gupta, A.K., Nadarajah, S.: Handbook of beta distribution and its applications. CRC press (2004)
9. Gyanchandani, M., Rana, J., Yadav, R.: Taxonomy of anomaly based intrusion detection system: a review. International Journal of Scientific and Research Publications **2**(12), 1–13 (2012)
10. Ma, Z., Leijon, A.: Beta mixture models and the application to image classification. In: Image Processing (ICIP), 2009 16th IEEE International Conference on, pp. 2045–2048. IEEE (2009)
11. Moustafa, N., Slay, J.: A hybrid feature selection for network intrusion detection systems: Central points (2015)
12. Moustafa, N., Slay, J.: The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems. In: Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), 2015 4th International Workshop on, pp. 25–31. IEEE (2015)
13. Moustafa, N., Slay, J.: Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). In: Military Communications and Information Systems Conference (MilCIS), 2015, pp. 1–6. IEEE (2015)
14. Moustafa, N., Slay, J.: The evaluation of network anomaly detection systems: Statistical analysis of the unsw-nb15 data set and the comparison with the kdd99 data set. Information Security Journal: A Global Perspective **25**(1-3), 18–31 (2016)
15. Moustafa, N., Slay, J., Creech, G.: Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks. IEEE Transactions on Big Data **PP**(99), 1–1 (2017). 10.1109/TBDATA.2017.2715166
16. Pontarelli, S., Bianchi, G., Teofili, S.: Traffic-aware design of a high-speed fpga network intrusion detection system. IEEE Transactions on Computers **62**(11), 2322–2334 (2013)
17. Rousseeuw, P.J., Hubert, M.: Robust statistics for outlier detection. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **1**(1), 73–79 (2011)
18. Saurabh, P., Verma, B.: An efficient proactive artificial immune system based anomaly detection and prevention system. Expert Systems with Applications **60**, 311–320 (2016)
19. Singh, K., Guntuku, S.C., Thakur, A., Hota, C.: Big data analytics framework for peer-to-peer botnet detection using random forests. Information Sciences **278**, 488–497 (2014)
20. Tan, Z., Jamdagni, A., He, X., Nanda, P., Liu, R.P.: A system for denial-of-service attack detection based on multivariate correlation analysis. IEEE transactions on parallel and distributed systems **25**(2), 447–456 (2014)