Reduction of the Feature Space for the Detection of Attacks of Wireles Sensor Networks

Victoria Korzhuk, Ilya Shilov, Julia Torshenko
ITMO University
Saint Petersburg, Russia
vika@cit.ifmo.ru, ilia.shilov@yandex.ru, ulka.torshenko@gmail.com

***Abstract*—The article evaluates the informativeness of the features of the abnormal behaviour of node in wireless sensor network. The estimation is carried out for the basic methods of attacking on wireless sensor networks, such as «funnel», «wormhole», «selective forwarding», etc. The estimation is performed using three basic methods: the method of Shannon, the method of Kullback and the method of accumulated frequencies. Special attention is paid to the dependence of the feature informativeness on various characteristics of the network (topology, packet generation periods, the degree of stochasticity of the selection of addresses for the generated packets transmission). Estimates are compared with previously obtained estimates for the simplest network with the mesh topology. Key results are the reduction of the feature space by uninformative features extracting (when reducing the introduced scale of feature informativeness degree is used), the formation of samples with estimates of informativeness for each network and each pair «normal behaviour»–«specific attack type». Also the program for automatic calculation of estimates of the informativeness and its subsequent analysis is created. In the future the obtained results can be used as the basis for methods of classification, aimed at identifying of anomalous behaviour in wireless sensor networks.**

I. Introduction

Wireless sensor networks, as the basis for the Internet of Things, represent a relatively new area of development of information technologies. Every year, the extent of use of such technologies increases; however, more and more information about the degree of success of its application appears. Much attention is paid to wireless sensor network security issues. Up to the present time many ways of attacks on wireless sensor networks have been described. The most common and widespread attacks described in the previous work.

One of the most commonly used approaches to information security of any information system (that uses or does not use wireless sensor network) is the creation of the intrusion detection system. In wireless sensor networks, this approach has not been applied so far: used protection methods are mostly symptomatic, i.e. are focused on solving a specific problem. At the same time fundamentally different methods of attack counteraction are used. These often contradict with each other. A vivid example is the use of the «wormhole» attack to protect against the «funnel» attack.

The main aim of the project is the creation of intrusion detection system using statistical methods. The first steps in this direction were made in the previous work (Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks). Here, the feature space was formalized and the first conclusions about the feasibility of this feature space usage were made. For this purpose three methods of estimating of the informativeness were used: the method of Shannon, the method of Kullback, and the method of cumulative frequencies. Also two important conclusions regarding the parameters of the sample were made:

1) Informativeness is higher when the statistics collecting period is longer;
2) Significant increase in the volume of sample leads to slight increase of feature informativeness. It should be noted that the sample size should be sufficient to meet the law of large numbers. In other words, the sample size has no effect on informativeness only if the sample is representative, i.e. reflects the real frequency distribution.

All conclusions were obtained using the model of attacks on wireless sensor networks described in previous work (The Model of the Attack Implementation on Wireless Sensor Networks). This paper provides further evaluation of the feature space in order to identify:

1) Always uninformative features. These can be discarded without harm to the classification carried out using the obtained feature space;
2) Features that are uninformative for specific characteristics of the network;
3) Dependencies between the features and characteristics of the network.

This article describes the same methods as in previous work (Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks). Re-description of methods of informativeness calculation is omitted for brevity.

II. Statement of the Problem

As it was already noted, the initial results of the evaluation of informativeness were given in the previous work. It should be noted that final conclusions about the feasibility of the use of certain features, as well as about the admissibility of exclusion of the part of the features from the sampling, can not only be made on the basis of these data. The main reason is the assumptions used in the process of estimation:

1) Used topology is the mesh network;
2) The period of packet generation in each node is the same and equal to 10 seconds; the beginning time for

different nodes is offset for uniformly distributed integer value in the range from 0 to 20;
3) The destination address of each packet is chosen randomly.

These assumptions with required accuracy interpolate behaviour of wireless sensor network, upon which there is a decentralized application operating. However, tree topology is used far more frequently nowadays. Furthermore, a major part of packets is delivered to a specific node (precisely, main coordinator of the network). The main reason is in youthfulness of wireless sensor networks: significant researches in this sphere are still in progress. Insufficient level of development leads to insufficient spread of decentralized applications built on top of wireless sensor networks, including ZigBee.

This paper covers following matters:

1) Information content evaluation for cluster tree topology and composition of ranked by information content list of features;
2) Information content evaluation in mesh network with determined packet routes and comparison of results with those acquired in previous work (Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks);
3) Information content evaluation in mesh network with diverse packet generation periods and comparison of results with those acquired in previous work (Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks);
4) Selection of least valuable features for various network characteristics and composition of not informative features list;
5) Formalization of common rules for selection of features when using data science algorithms.

Evaluation is performed on 15 classes. Classes are listed in Table I.

TABLE I. REVIEWED TYPES OF SYSTEM BEHAVIOR

Class	Description
denial_of_sleep	Denial of sleep attack. Attacker generates packets dedicated to one precise node in the network. In case of determined routing this node is different from the one to which major part of packets is usually delivered, as attack if nonsense otherwise.
flood	Flood – packet generation with high rate for precise node or set of nodes.
normal	Normal system behaviour.
repeated_transmission	Repeated transmission – every N-th packet is stored in inner queue. Repeated transmission (if packets to transmit are present) is performed every K seconds.
repeated_transmission_dest	Repeated transmission of packets for node A - every N-th packet for node A is stored in inner queue. Repeated transmission (if packets to transmit are present) is performed every K seconds.
repeated_transmission_src	Repeated transmission of packets from node A - every N-th packet from node A is stored in inner queue. Repeated transmission (if packets to transmit are present) is performed every K seconds.
selective_forward	Selective forward – every N-th packet is discarded.
selective_forward_dest	Selective forward for node A – every N-th packet for node A is discarded.
selective_forward_src	Selective forward from node A – every N-th packet from node A is discarded.
sinkhole	Sinkhole attack – described in previous work
spoof	Spoofing – for every N-th packet created by attacker both source and destination addresses are chosen randomly.
spoof_dest	Spoofing – for every N-th packet created by attacker source address is chosen randomly and destination addresses is always the same.
spoof_src	Spoofing – for every N-th packet created by attacker destination address is chosen randomly and source addresses is always the same.
sybil	Sybil attack – described in previous work
wormhole	Wormhole attack – described in previous work

III. RANKING INFORMATION CONTENTS

For simplification sake let us construct ranks for features:

1) HI – High Information Content;
2) UI – Upper Information Content;
3) MI – Medium Information Content;
4) LI – Lower Information Content;
5) NI – Negligible Information Content.

In Table II rules for translation from information content evaluation scales to constructed scale are given.

The main purpose of this scale is rationale for setting borders when truncating feature set and selecting most informative features.

TABLE II. TRANSLATION INTO INFORMATIVITY SCALE

Scale	Shannon's method	Kullback's method	Accumulated frequencies method
HI	(0,8; 1,0]	$\left(x_{min} + 4 * \frac{x_{max} - x_{min}}{5}; x_{max}\right)$	
UI	(0,6; 0,8]	$\left(x_{min} + 3 * \frac{x_{max} - x_{min}}{5}; x_{min} + 4 * \frac{x_{max} - x_{min}}{5}\right)$	
MI	(0,4; 0,6]	$\left(x_{min} + 2 * \frac{x_{max} - x_{min}}{5}; x_{min} + 3 * \frac{x_{max} - x_{min}}{5}\right)$	
LI	(0,2; 0,4]	$\left(x_{min} + \frac{x_{max} - x_{min}}{5}; x_{min} + 2 * \frac{x_{max} - x_{min}}{5}\right)$	
NI	(0; 0,2]	$\left(x_{min}; x_{min} + \frac{x_{max} - x_{min}}{5}\right)$	

IV. INFORMATION CONTENT IN CLUSTER TREE

The opportunity of constructing cluster tree in network with ZigBee stack is caused by usage of IEEE 802.15.4 as physical and MAC standard, since it defines such a structure. Experiment considers network of 15 nodes, connections between nodes are pictured in Fig. 1.

First of all during the experiment three sets for different periods of statistics collection were acquired: 10*T, 60*T, 360*T. At Fig. 2 evaluation of information content (Shannon's method) is given. These results are actually similar to those obtained in work "Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks".

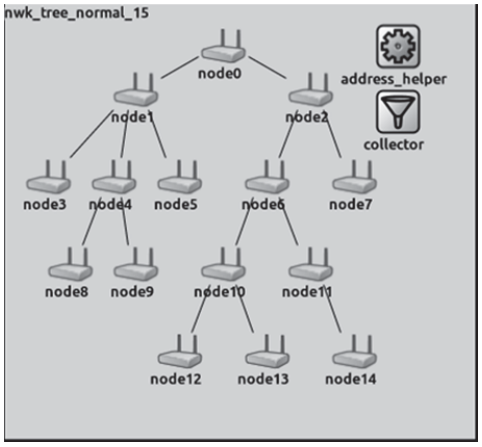

Fig. 1. Cluster Tree topology

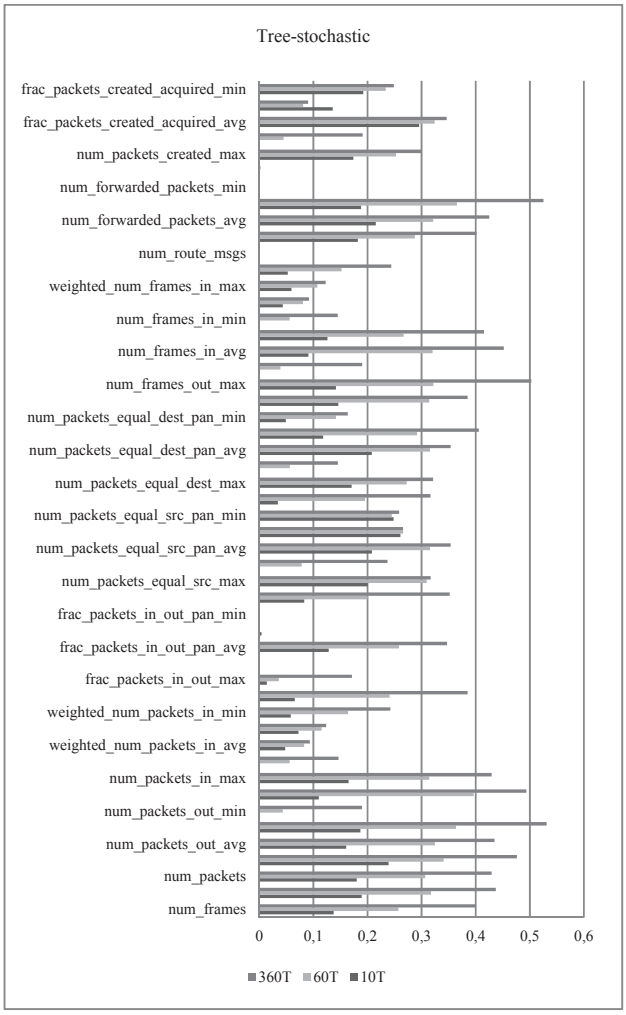

Fig. 2. Information content (Shannon's method) for cluster tree

Then information content assessments for two-class classification were calculated – for every pair "normal behaviour–attack". Tables III-V present five most informative features of denial of sleep attack for every period. All evaluations were obtained with three methods of feature evaluation. Rest 15 sets are not given explicitly due to their extents.

Like in mesh network a tendency to overall information content growth with period incrimination is observed. Besides, again existence of features appropriate for abnormal behaviour detection is proven. Yet before data mining algorithms usage it is difficult to say, whether these features are able to reveal exact attack type, it is possible to assume that high accuracy will be achieved with the help of boosting algorithms, i.e. algorithmic compositions – in particular logical classifiers, accurate only on subsets of data.

For comparison of most informative features for different topologies (mesh and cluster tree) maximum period of statistics collection was chosen – 360T. At Fig. 3 most informative features are depicted.

TABLE III. 5 MOST INFORMATIVE FEATURES WITH SHANNON'S METHOD

Period	Feature	Information content
10T	num_packets_equal_src_max	1
	num_packets_equal_dest_max	1
	num_packets_created_max	1
	num_packets_equal_dest_pan_max	0.388859
	num_packets_in_max	0.101423
60T	num_packets_equal_src_max	1
	num_packets_equal_dest_max	1
	num_packets_created_max	1
	num_packets_equal_dest_pan_max	0.995858
	frac_packets_in_out_avg	0.549236
360T	num_packets_equal_src_avg	1
	num_packets_equal_src_max	1
	num_packets_equal_dest_max	1
	num_packets_equal_dest_pan_max	1
	num_packets_created_max	1

TABLE IV. 5 MOST INFORMATIVE FEATURES WITH KULLBACK'S METHOD

Period	Feature	Information content
10T	num_packets_equal_dest_pan_max	2.277001901
	num_packets	0.867413464
	num_packets_in_max	0.85058403
	frac_packets_in_out_avg	0.786712173
	num_packets_avg	0.739905533
60T	frac_packets_in_out_avg	5.286882719
	frac_packets_in_out_pan_avg	4.510871542
	num_packets	3.857940948
	num_packets_in_max	3.316789377
	num_packets_avg	3.200612921
360T	num_packets	13.49451084
	num_forwarded_packets_max	11.24762801
	num_frames	11.08580809
	num_packets_out_max	10.57995196
	num_frames_avg	10.36466165

Since even maximum values of information content do not reach HO rank, features with information content higher than NI are shown (important: here and later on all features that are not NI for at least one data set are chosen for plot).

Following conclusions can be made while analysing the results:

1) In general information content in cluster tree is higher than in mesh network;

2) For both cluster tree and mesh most informative features are the same, yet their order in sorted list may vary.

TABLE V. 5 MOST INFORMATIVE FEATURES WITH ACCUMULATED FREQUENCIES METHOD

Period	Feature	Information content
10T	num_packets_equal_src_max	500
	num_packets_equal_dest_max	500
	num_packets_created_max	500
	num_packets_equal_dest_pan_max	316
	num_packets	165
60T	num_packets_equal_src_max	500
	num_packets_equal_dest_max	500
	num_packets_created_max	500
	num_packets_equal_dest_pan_max	499
	frac_packets_in_out_avg	390
360T	num_packets_equal_src_avg	167
	num_packets_equal_src_max	167
	num_packets_equal_dest_max	167
	num_packets_equal_dest_pan_max	167
	num_packets_created_max	167

V. INFORMATION CONTENT AND NETWORK CHARACTERITICS

Second experiment is directed to evaluation of correlation between information content of features and network characteristics:

1) Packet generation periods;
2) Level of randomity when selecting destination addresses.

The first problem is solved with appointment of own values of expected value for Gaussian distributions, determining package generation periods. At Fig. 4 mesh network is depicted. Expectations in first case are the same (10) – this case is considered in previous work (Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks) – in second case these are set arbitrarily (values are given in Table VI). Worth mentioning: every node is actually PAN of 5 nodes. Thus every real node is in fact generating new packet every 5*T seconds. Number of nodes is chosen arbitrarily and is not significant as results for different values could be reduced to the results of this work by changing expectation.

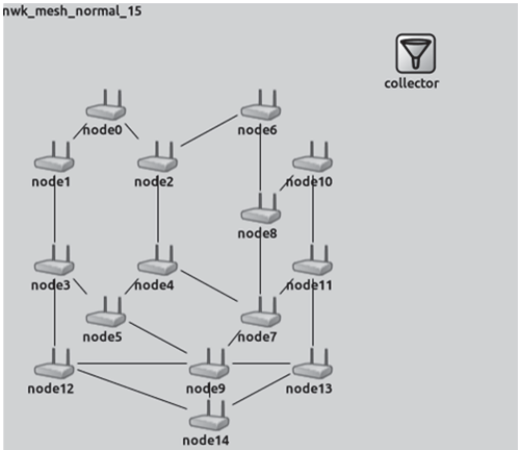

Fig. 4. Mesh network

Fig. 3. Informative features for mesh and cluster tree

Second part of the experiment is following: every node (except for coordinator) is given addresses, which are used as destination addresses for all generated packets. Routes are manually created to make attacks like repeated and selective forwarding and sinkhole having meaning. Table VII describes destination addresses for every node in the network.

TABLE VI. EXPECTATIONS FOR GAUSSIAN DISTRIBUTION

Node	Expectation	Node	Expectation
Node 0	5	Node 8	15
Node 1	10	Node 9	20
Node 2	15	Node 10	10
Node 3	10	Node 11	15
Node 4	20	Node 12	10
Node 5	15	Node 13	10
Node 6	5	Node 14	10
Node 7	20		

Modelling of *selective_forward_dest* and *repeated_transmission_dest* is not performed in case of determined routing as in this case such attacks replicate *repeated_transmission* and *selective_forward*.

TABLE VII. DETERMINED ROUTES FOR MESH NETWORK

Node	Destination node	Node	Destination node
Node 0	Random	Node 8	Node 10
Node 1	Node 0	Node 9	Node 12
Node 2	Node 0	Node 10	Node 14
Node 3	Node 12	Node 11	Node 10
Node 4	Node 0	Node 12	Node 14
Node 5	Node 0	Node 13	Node 12
Node 6	Node 10	Node 14	Node 0
Node 7	Node 11		

Generalized results of the experiment are given at Fig 5. Like in case of mesh and cluster tree comparison all the features with rank higher than NI for at least one set are given. Worth mentioning that evaluations for each pair "normal-abnormal behaviour" were also obtained (again with three methods of evaluation). In this paper these results are not given due to their extent. General conclusion is the same as for similar evaluation for cluster tree: for most attacks absolutely informative features exist, which theoretically causes high accuracy of boosting methods.

Following conclusions come after analysis of given information:

1) Information content of features in network with determined routes is usually higher than in stochastic network;
2) Number of informative features in network with determined routes is usually higher than that in stochastic network;
3) Information content weakly correlates with fraction of packet generation periods by different nodes, which again stresses justice of given in the work "The Model of the Attack Implementation on Wireless Sensor Networks" formula for calculation of total packet generation frequency. Reverse value – average package generation period – has exceptional significance, since it partially defines statistics collection period.

VI. DISCRETIZATION PARAMETER SELECTION

Another variable connected with information content is discretization parameter. As noted in work "Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks", information content evaluation methods operate with finite sets of values. To perform translation into discrete scale of finite number of values following actions were committed:

1) Search for maximum and minimum elements;
2) Calculation of discretization step as quotient from difference between maximum and minimum elements and discretization parameter.

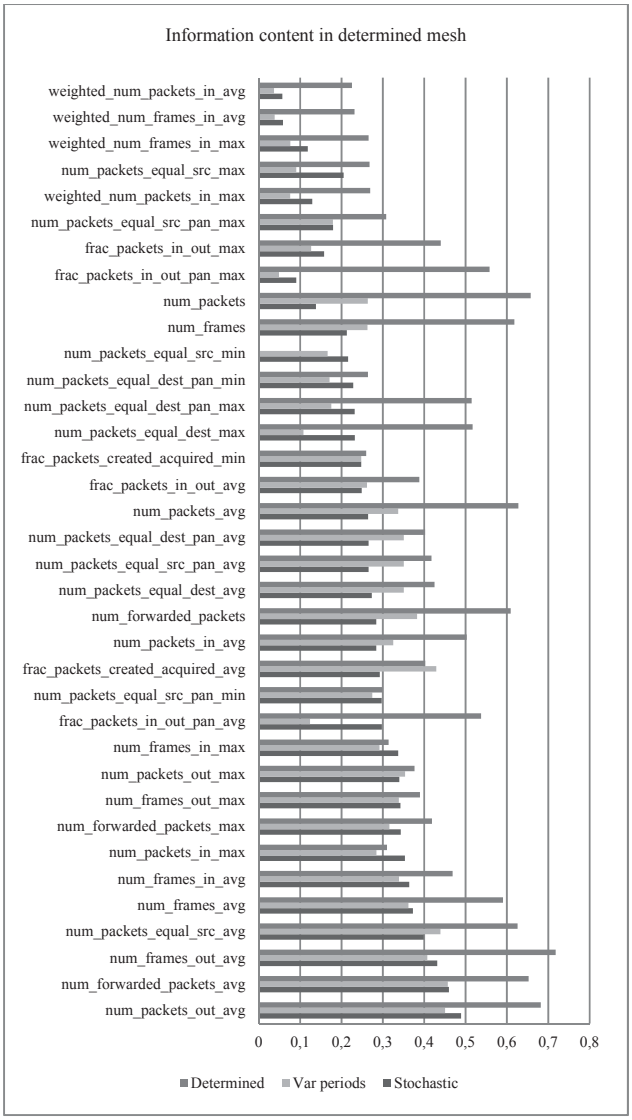

Fig. 5. Information content in mesh network

Paper "Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks" and this paper use discretization parameter of value 10. Worth mentioning that in some cases this parameter can have substantial meaning. For instance, if values of feature vary within limits less than chosen discretization parameter (maximum and minimum values are blowouts), than even if every class is given finite set of feature values and the sets do not cross, calculated information content is about to be zero: all values will be translated to the same one.

For analysis of correlation between information content and discretization parameter and evaluation of all not NI features for two topologies was performed (statistics collection period – 360*T). Results are given at Fig. 6-7. General conclusions:

1) The greater discretization parameter is, the greater information content evaluations are;
2) The greater discretization parameter is, the more features are not NI.

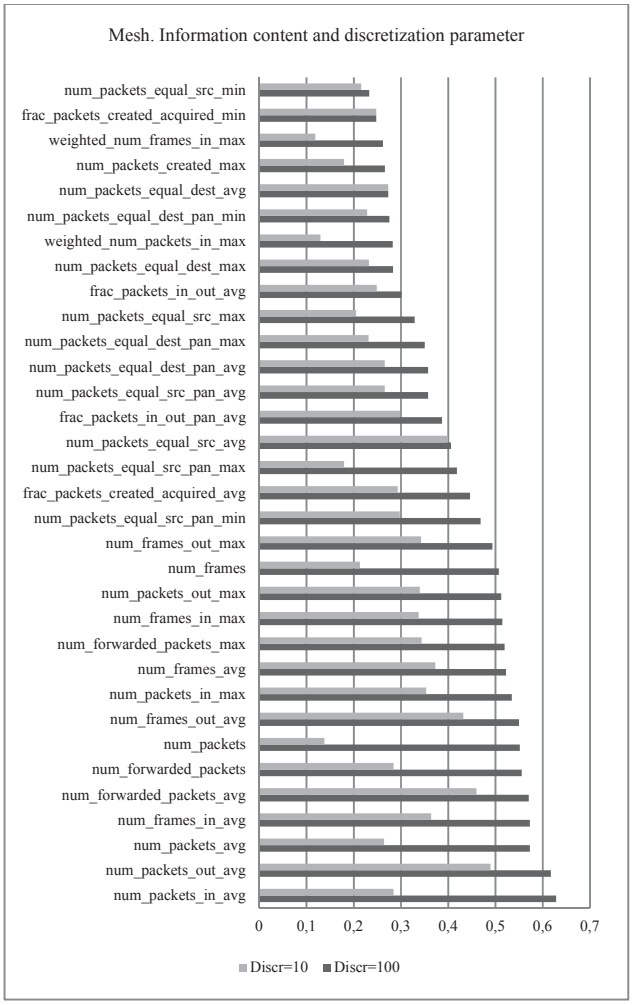

Fig. 6. Information content and discretization parameter in mesh network

VII. Discretization parameter Selection

The last result shown in this work is amputation of not informative features. At the step of feature set formalization the redundancy of several features was out of concern. As a result, the power of feature set was more than 50. In practice, such number of features is rarely used. For instance, in logical methods of classification it's not recommended to use more than 7-10 features.

Thus, when possible redundant features are to be deleted. As a rationale in this paper all obtained information content evaluations are used – for all types of networks:

1) Mesh network:

a) Stochastic, with equal expectations of Gaussian distributions defining periods of packet generation;
b) Stochastic, with different expectations of Gaussian distributions defining periods of packet generation;
c) Determined.

2) Cluster tree.

On basis of evaluations, acquired with Shannon's method, it is possible to perform the reduction of feature set. For that it's necessary to use values of information content for every feature obtained for different networks. It was decided to delete following features:

1. For all networks:
   1.1. num_packets_out_min;
   1.2. num_packets_in_min;
   1.3. weighted_num_packets_in_avg;
   1.4. frac_packets_in_out_min;
   1.5. frac_packets_in_out_pan_min;
   1.6. frac_packets_in_out_pan_max;
   1.7. num_packets_equal_dest_min;
   1.8. num_frames_out_min;
   1.9. num_frames_in_min;
   1.10. weighted_num_frames_in_avg;
   1.11. num_route_msgs;
   1.12. num_packets_created_avg;
   1.13. num_packets_created_min;
   1.14. frac_packets_created_acquired_max;
2. Additionally for mesh:
   2.1. weighted_num_packets_in_min;
   2.2. weighted_num_frames_in_min;
3. Additionally for cluster tree:
   3.1. frac_packets_in_out_max.

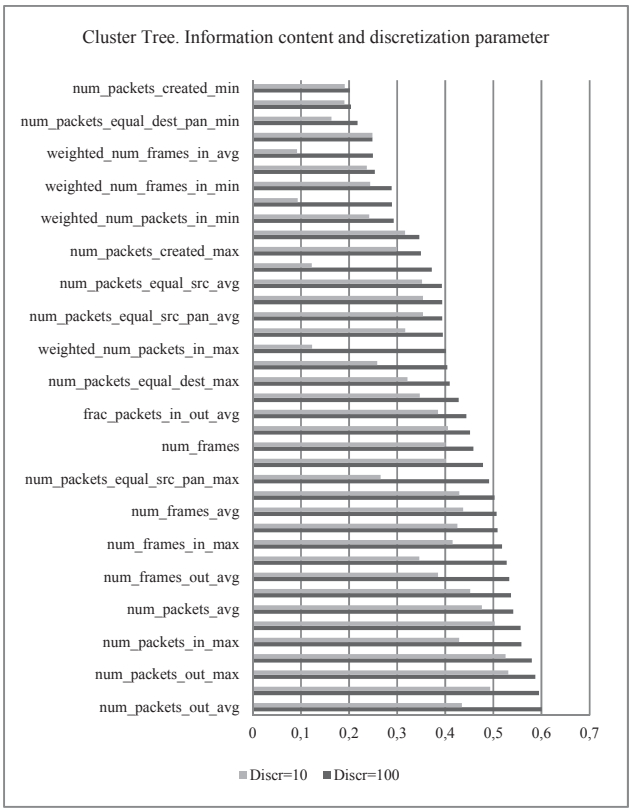

Fig. 7. Information content and discretization parameter in cluster tree

Important remarks concerning usage of feature set in data mining are to be made. During reduction, only NI features were deleted. Thus, the obtained feature set is about to be reduced even more. Most attention should be paid to two factors:

1) Network characteristics:

a) Topology;
b) Average packet generation period;
2) Level of "similarity" of features.

By "similarity" we assume following: difference between some features is insignificant. For instance, in the set of features *num_packets* and *num_frames* are present. The second feature was introduced to catch the correlation of information content and packet size without usage of byte-values. If both features are informative, it's reasonable to leave only one (*num_packets* is better as is has lower values and reduces stochastics).

Furthermore, the selection of features is necessary not for all data mining algorithms. For example, it's necessary for logical classification. In case of usage of linear methods due to regularization selection of features is performed automatically. However, linear methods perform only two-class classification. Thus, when using these methods, it's necessary to solve classification problem several times:

1) To revel possibility of every attack in case of classification "normal – abnormal behaviour" for every attack;
2) To identify exact type of attack – classification "attack_1 – attack_2".

In all cases the evaluation of information content should be performed multiple times. To automate process of information content evaluation a script in Python programming language was written. The functional of the script:

1) Calculation of information content with Shannon's method for N classes;
2) Calculation of information content with Shannon's and Kullback's method and with method of accumulated frequencies for every pair of classes in set of classes of power N;
3) Deletion not informative features;
4) Sorting of features by information content values;
5) Selection of most informative features for K of N classes.

VIII. Conclusion

From point of view of data mining and applying classification algorithms it's vital to define feature set and select most informative features. The paper continues examination of formalized in [2] feature set. The main attention is paid to investigation of correlation between information content and different network characteristics – mostly with topology and packet generation periods. The main result is reduction of feature set for all possible wireless sensor networks built with ZigBee stack of protocols. Theoretically all the conclusions are valid for other wireless (and even wired) networks – in case of adequate choice of parameters for model [1]. In future obtained results will be used as a basis for classification with the help of various methods of data mining: logical, metric, linear, stochastic and others.

References

[1] D.E. Comer, Internetworking With TCP/IP Vol I: Principles, Protocols, and Architecture. Pearson, 2014.
[2] I.S.Lebedev, I.E.Krivtsova, V.Korzhuk, N. Bazhayev, M.E. Sukhoparov, S. Pecherkin, K. Salakhutdinova, "The Analysis of Abnormal Behavior of the System Local Segment on the Basis of Statistical Data Obtained from the Network Infrastructure Monitoring", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) - 2016*, Vol. 9870, pp. 503-511
[3] A.S. Tanenbaum and D.J. Wetherall, *Computer Networks*. Prentice Hall, 2011
[4] P. Baronti, P. Pillai, V.W.C. Chook, S. Chessa, A. Gotta, Y. Fun Hu, *Wireless sensor networks: A survey on the state of the art and the 802.15.4 and ZigBee standards*, "Computer Communications", vol. 30, Dec. 2007, pp.1655–1695.
[5] F. Cuomo, S. Della Luna, E. Cipollone, P. Todorova, T. Suihko, *Topology Formation in IEEE 802.15.4: Cluster-Tree Characterization*, in Proc. PerCom Conf., March 2008, pp.276-281.
[6] F. Cuomo, E. Cipollone, A. Abbagnale, Performance analysis of IEEE 802.15. 4 wireless sensor networks: An insight into the topology formation process, "Computer Networks", vol.53, Dec. 2009, pp. 3057-3075.
[7] A.M.Wyglinski, K.Huang, T. Padir, L. Lai, R.T.Eisenbarth, and K.Venkatasubramanian *Security of Autonomous systems using embedded computing and sensors,* "IEEE micro 33 (1) 2013", art. no. 6504448, pp. 80-86
[8] C. E. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J., vol. 27, 1948, pp. 379-423.
[9] S. Kullback, *Information Theory and Statistics*. Peter Smith, 1978.
[10] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[11] I.S.Lebedev, V.Korzhuk, I.Krivtsova, K.Salakhutdinova, M.E.Sukhoparov, D.Tikhonov, *Using Preventive Measures for the Purpose of Assuring Information Security of Wireless Communication Channels,* "Proceedings of the 18th Conference of Open Innovations Association FRUCT" - 2016, pp. 167-173
[12] S.Y. Novak, Extreme Value Methods with Applications to Finance. CRC Press, 2012.