Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks

Igor A. Zikratov, Victoria Korzhuk, Ilya Shilov, Alexey Gvozdev
ITMO University
Saint Petersburg, Russia
{zikratov, vika, a.gvozdev}@cit.ifmo.ru, ilia.shilov@yandex.ru

Abstract—**The article describes the formalization of the feature space in order to detect abnormal behaviour of nodes in wireless sensor network using statistical methods. The main methods of destructive impact on the infrastructure of wireless sensor networks based on ZigBee Protocol stack are considered. Special attention is paid to attacks on integrity and availability, which theoretically can be detected using the methods of machine learning and mathematical statistics. On the basis of standards and specifications, as well as considered attacks, the space of more than 50 features is developed. Using the methods of Shannon, Kullback and accumulated frequencies, informative value of formalized signs was evaluated. Conclusions about the existing dependencies between the information content of features, the statistics collection period and sample size used to calculate the information content are drawn. Received the results can be used as a basis for further evaluation of the most suitable characteristics for the classification of attacks depending on the network characteristics. In the future the main aim of the study is to build an intrusion detection system that uses statistics of the interactions for a certain period of time as a source of information about the system.**

I. Introduction

Wireless sensor networks are the networks containing small devices equipped with sensors and using wireless technologies for information transmission. The main features of the device are the relatively low data transmission rate and energy savings.

The great number of attacks on these systems was described earlier. With the advent of each new attack, the basic methods of countering are being formalized. These often depend on the method of the attack (i.e., are symptomatic): some attacks can be prevented by cryptographic methods of protection, others – with the use of filtration, others - through methods of integrity monitoring, and so on.

From the point of view of wireless sensor network security, like any other technology, the most important task is to identify anomalous network behaviour for later identification of conducted attack and for application of an adequate method of resistance. One of the possible ways of assessing anomalous network behaviour is to analyse frequency characteristics of the network obtained for the certain period of time.

In the previous work the model of carrying out the attacks on wireless sensor networks was described. Using the software implementation of this model it is possible to gather statistical information about the system operation and functioning in different modes: in the normal state and while the certain attack implementation.

The aim of this work is the formalization of the feature space and its further reduction via the various criteria of informativeness. Further, features would be used for the formation of the «objects-features» matrix and for the teaching different machine learning algorithms on the basis of this matrix. The resulting algorithm works on real data and, using the user-defined (or algorithm-defined) time intervals, estimates the «normality» of the system behaviour over the past period of time.

II. Attacks on wireless sensor networks

Traditionally it is accepted to divide all existing attack to classes according to the properties of information security: confidentiality, integrity and availability.

Attacks on confidentiality

The attack on the confidentiality of wireless sensor networks is reduced to the attack on the confidentiality of sent messages. If the traffic is not encrypted, than the aim of the attack is simply listening to network traffic and its subsequent analysis. Protection from interception is realized solely by technical means of information protection. The obvious solution is to encrypt transmitted traffic. ZigBee and IEEE 802.15.4 uses AES-128. With the use of protective encryption, the attack on privacy is either the attack on the encryption algorithm, or unauthorized physical access to the device of information transmission, i.e. to the router (or the coordinator, which in this case does not matter).

Attacks on the availability

Currently, there are four basic ways to carry out attacks on availability in wireless sensor networks:

1) Denial of sleep;
2) Flood;
3) Sinkhole attack;
4) Sybil attack.

Denial of sleep is the attack that is specific to wireless sensor networks. As noted previously, a significant consideration while functioning and operating protocol designing of wireless sensor networks is paid to the issue of energy efficiency. In ZigBee, as in IEEE 802.15.4, all devices are divided into three classes:

1) Coordinator;
2) Routers;
3) End devices.

Routers usually play the role of coordinators for PAN (Private Area Network). Each PAN coordinator is the only one. The router, which is the coordinator in some PAN, can be a subordinate device in another PAN. There is two-operation mode of the coordinator (and therefore of the network): with beacon sending and without beacon sending. In the first mode, the coordinator periodically sends a broadcast frame of channel layer (beacon). It contains information about the network configuration. The beacon also performs the functions of synchronization: after the beacon there is active period that is divided into 16 time slots (the first slot is the beginning of the transmission of the first data bit of the beacon). Two periods are built on time slots: CAP (Content Access Period) and CFP (Content Free Period). During the first period there is a competition for the communication environment according to the algorithm CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance). During the second period (which may not be present) nodes provide guaranteed information transmission (using the concept of GTS – Guaranteed Time Slots). After the active period there is an inactive period (period of low power or a sleep period) when devices do not interact with the environment. In the network mode without beacon sending the main coordinator and routers do not go to sleep mode.

Routers store information intended for end devices that spend the most of time in low power mode. Information about data availability for the end devices is transmitted in the beacon, which, in the case of the network without beacons, is requested explicitly. If there is no data to receive and the end device does not need to transmit information to the coordinator, the end device goes into the low power mode until the inactive period begins. To obtain the data, the end device sends to the coordinator a dedicated POLL request. Information transmission by coordinator is either in CAP (in condition of generic competition), or CFP. All this time, the end device is working in normal mode of power consumption. The more data is received by the destination device, the less the inactive period is and the faster the battery will run down. Using the «Denial of sleep» attacker tries to achieve this purpose: the number of sent packets is larger than the device usually receives.

Flood is an attack that is typical for almost every network. It actually consists in the flooding of a certain area of the network by the large number of packets. It causes significant decrease in productivity. The attacking node sends the large number of packets to the certain address of the node or PAN.

Sinkhole is observed if one or several nodes concentrate the larger part of the network routes. The packet transmission route is understood as the sequence of hops (connections between adjacent nodes) that the packet passes during transmission from the source to the recipient. The route is determined by the internal tables of nodes, which are constructed by the AODV (Ad-hoc On Demand Vector algorithm) algorithm. Each record in the table matches some destination address in the network and contains the address of the next node in the route. In the absence of the required recording, route information is obtained using a broadcast request. Each node received this request sends it back via broadcast, if it is not the destination node. Protection from re-transmission is carried out by the counter of passed hops in each broadcast packet. When the request is received, the destination node sends a response back to the source through the known route computed using information from the request packet and internal tables of the nodes. In the case of «funnel» attack the most of these routes will pass through one or several nodes that are able to carry out attacks, for example, on the integrity of transmitted packets.

Sybil attack is typical for peer-to-peer networks. There are two options to carry out this attack. First, attacking units can «capture» all connections of the node or group of nodes and, thus, isolate them from the external network, filtering or not passing packets. Secondly, the node in the network can get the «multiple personality», i.e. to act as multiple nodes. In this case, the attack can significantly reduce the network performance: the attacking node holds the forwarded packets because of a large number of nodes, which the packet must go through. Also there is meaningless packet transmission to the channel: such packets are destined for nodes that do not really exist.

Attacks on the integrity

There are two groups among the attacks on the integrity in wireless sensor networks. The first group includes the attacks on the semantic integrity of the forwarded packets. Protection from this attack was the use of checksums, including ones based on a cryptographically strong hash algorithms and the computation of the digital signature. Therefore, a simple change of the content of the packet is monitored by checking the checksum matching: one is specified in the packet and other is calculated when the packet was received. Therefore, the attack on the integrity of this group is currently reduced to attacks on algorithms for checksum computing and, as a result, on the hashing algorithms. In cryptography, such problems are called the search of collisions of the first and second kind. The collision of the first kind implies the existence of an effective way of finding two messages with the same hash value. Collision of the second kind implies the existence of an algorithm that can construct a new message with the same hash value using a known message and the hash.

The second group consists of attacks that can be tracked statistically. These include:

1) Selective forwarding;
2) Spoofing;
3) Re-transmission;
4) Wormhole.

Selective forwarding is a traditional type of attack of the network layer connected with routing. In complex network topologies, such as mesh network and cluster tree, before reaching the destination the packet passes through a sequence of intermediate routers. Each of the routers can filter passing packets and discard certain packets. Therefore, this attack may be interpreted not only as the attack on the integrity (in this case, the integrity of the transmission path of the message), but as the attack on availability. Most often, the attacker still sends a portion of the packet to avoid disconnection from the target node due to the reaction of existing methods of protection.

Spoofing is an attack, when attacking node sends packets in which the source is indicated as other node. In essence this attack in networks is analogous to identity fraud. Existing methods of protection are reduced to the application of digital signatures of messages that does not always work, because of the attacker can use previously held packets, although this is a different kind of attack. In addition, the problem of key information preserving still exists. Usually the end devices, as noted earlier, have low memory and low processing power of computing modules. So, the keys are often stored either in the memory of routers, or in the memory of allocated device, which is the key information server. In this case, any method of unauthorized access to the server leads to compromising of the keys. However, statistically spoofing can still be tracked.

While *re-transmission*, the malicious node performs saving of certain messages to be forwarded. Then packets can be forwarded again. Potential methods of protection are the use of timestamps and packet counters. The disadvantage of the first method is that re-transmission can be carried out in the period when the timestamp is still considered to be true. In this case, for example, the action called when receiving a packet can be called multiple times due to receiving multiple packets. The use of the counter is associated, firstly, with the need to store information about the values from the counters in all the latest packets that were received from each node in the network, and secondly, with the need of data synchronization between different nodes in the initial phase of work. In addition, the counter value can be spoofed by the attacker. The result is the usage of digital signature, which is not always a panacea, as noted earlier.

Wormhole is a method of attack based on the use of two or more associated by high-speed connection attacking nodes located at a considerable distance from each other. There are two possible effects of such compounds. First, the node performing forwarding of passing packets can primarily send the packet to the node associated with the wormhole. Then the packet either with changes or no changes is sent to the network and can reach the destination earlier than the original packet passing along the true route. In some cases this allows to bypass the timestamps security, and sometimes to cause abnormal sequence of actions: by and large, the system re-sends the message when a duplicate reaches the destination earlier than the original.

The second method of attack is associated with the traffic pass through the high-speed channel. In this case, the nodes connected by the wormhole, use each other when routing as the neighbouring nodes. As a result, if there are a lot of attacking nodes or network graph is divided into several strong components connected by bridges, than the significant part of the routes would pass through the wormhole. In this case, the «wormhole» attack turns to «funnel» attack.

Methods of protection

To ensure the confidentiality and integrity different cryptographic methods of protection are applied: encryption, hashing, digital signature, exchange of key information, etc. Physical security of wireless sensor network units is provided by either technical solutions or organizational measures. The interception of transmitted data through the environment is suppressed by technical measures of protection such as those that used for protection from side electromagnetic radiation and interference. It should be noted that the use of such protection methods in domestic wireless sensor network is rarely economically justified.

As it was noted earlier, many existing methods of protection against attacks on wireless sensor networks are symptomatic or problem-oriented, i.e., directed onto the specific threat repulse. The example is the «forced sleep»: the device is forcibly switched into energy saving mode if the time of active mode greatly exceeds the permissible thresholds.

Sometimes the behaviour that in other cases can be regarded as abnormal is used for protection. For example, in order to protect from flooding and «funnel» attack it can be applied the deliberate connection of some nodes located far from each other via the high-speed channel, that is nothing like the wormhole. For protection against wormholes the node geographic location sending in the packets can be used.

Statistical methods are also used, although these are extremely primitive: a periodic check of some network characteristics is embedded in program code in terms of exceeding user-defined thresholds. When detecting a potential attack, node automatically rebuilds the routes or even detaches from the network, and performs reconnection.

The alternative approach investigated in the work involves the collection of statistical information about the whole network interactions. This allows the use of the unified method for detecting of attacks on the wireless sensor network. It should be noted that the elimination patterns manually is not possible because of the enormous stochasticity of the processes occurring in wireless sensor networks.

Therefore, the task to be solve firstly is to select the most informative features from the point of view of detection of attacks on wireless sensor network and to compare machine learning methods that solve the classification problem from the point of view of accuracy of definition of abnormal behaviour based on the selection of features and from the viewpoint of practical applicability.

In practice, the obtained intrusion detection system aims to be trained on real data or on data close to real and received using network model from the previous work. Theoretically, reinforcement learning is possible when the classification algorithm receives additional configuration based on data received after the beginning of exploitation of the intrusion detection system.

III. The Formalization of the Feature Space

The significant problem in the formalization of the feature space is definition of method of data collecting, in this case, the statistical data. Considering the system that works in normal operation or under attack. After some equal time periods, statistics about the system during the last period is recorded to a file. The choice of recording period is one of the problems solved in this work.

While studying the IEEE 802.15.4 and ZigBee specifications, practical using of the system and analysing of existing ways of committing attacks on wireless sensor network based on ZigBee Protocol stack, more than 50 potential features of anomalous behaviour in the network were allocated. All features can be divided into several groups:

1) Quantitative features are presented in Table I;
2) Aggregate features, consisting of three values: maximum, minimum and average, are presented in Table II (for all the features three values are collecting: max - maximum, min - minimum, avg - averaged over the number of nodes in the network);
3) The features-ratios are presented in Table III (for all the features the three values are collecting: max - maximum, min - minimum, avg – average number of nodes in the network; if one (any) of the elements of the ratio is equal to zero, the ratio is equal to zero too, regardless of which part of the ratio was zero (numerator or denominator)).

TABLE I. Quantitative features of abnormal behaviour in wireless sensor networks.

Feature	Description
num_frames	Total number of frames transmitted in the network according to IEEE standard 802.15.4
num_frames_avg	Total number of frames transmitted in the network according to IEEE 802.15.4, averaged over the number of PAN in the network
num_packets	Total number of packets transmitted in the network according to the ZigBee specification
num_packets_avg	Total number of packets transmitted in the network according to the ZigBee specification, averaged over the number of PAN in the network
num_route_msgs	The number of transmitted routing messages (RREQ, RREP) according to IEEE standard 802.15.4
num_forwarded_packets	Total number of messages transmitted in the network within the packet routing protocol

It should be noted that in some networks num_packets may be a linear combination of num_forwarded_packets and num_packets_created. In this case, special attention must be paid to the choice of the method of machine learning.

In the case of linear machine learning methods it is necessary to avoid linear dependences between the features, although in some cases the problem is solved by regularization.

IV. The Methods Used for Informativeness Estimation

Before training and comparison of classification algorithms it is necessary to evaluate the selected features from the point of view of criteria of informativeness and to discard the least informative features. The informativeness in this case is understood as the value that becomes greater when the feature divides the sample into classes more and more accurate. In other words, feature will be more informative while it takes different values for different classes of objects and the same values for objects of the same class.

Three methods of informativeness assessment are used: the method of Shannon, the Kullback method and the method of cumulative frequencies.

TABLE II. The aggregated feature of abnormal behavior in wireless sensor networks

Feature	Description
num_packets_out	The number of packets sent by each node (own and forwarded)
num_packets_in	The number of packets received by each node (addressed to a node and to be forwarded)
weighted_num_packets_in	The number of packets weighted by the number of recipient nodes received by each node (addressed to a node and to be forwarded)
num_packets_equal_src	The number of received packets in which the sender is the same node
num_packets_equal_src_pan	The number of received packets in which a PAN sender it the same PAN
num_packets_equal_dest	The number of received packets in which the destination node is the same node
num_packets_equal_dest_pan	The number of received packets in which as the PAN recipient is the same PAN
num_frames_out	The number of frames sent by each node (own and forwarded)
num_frames_in	The number of frames received by each node (addressed to a node and to be forwarded)
weighted_num_frames_in	Взвешенное по числу узлов-получателей количество кадров, полученных каждым узлом (адресованных узлу и подлежащих пересылке) The number of frames received by each node (addressed to a node and to be forwarded) weighted by the number of recipient nodes
num_forwarded_packets	The number of forwarded packets by node
num_packets_created	The number of packets created by the node

TABLE III. Features-correlators for abnormal behavior detection in wireless sensor networks

Feature	Description
frac_packets_in_out	Ratio of the number of received packets and packets transmitted into the network for each node
frac_packets_in_out_pan	Ratio of the number of received and transmitted packets for each PAN
frac_packets_created_acquired	Ration of the number of packets created by the node and number of received packets in which the source is specified as this node

The method of Shannon. This method evaluates the informativeness from the point of view of information theory as the average amount of information (knowledge uncertainty reduction), attributable to different gradations of a feature. The informativeness of the feature is estimated as follows:

$$I(x) = 1 + \sum_{i=1}^{G}\left(P_i * \sum_{k=1}^{K} P_{i,k} * \log_K P_{i,k}\right) \tag{1}$$

$$P_i = \frac{\sum_{i=1}^{K} m_{i,k}}{N} \tag{2}$$

$$P_{i,k} = \frac{m_{i,k}}{\sum_{i=1}^{K} m_{i,k}} \tag{3}$$

Where:

G is the number of gradations of the feature *x*;

K is the number of classes;

N is the number of objects of all classes;

$m_{i,k}$ is the number of objects of class k, where the feature takes the value of gradation *i*;

P_i is the frequency of occurrence of gradation *i* to all objects of the sample;

$P_{i,k}$ is the proportion of objects of class *k* among all the objects for which the feature takes the value of gradation *i*.

The significant advantages of the method of Shannon are:

1) The possibility of informativeness estimation for several classes;
2) The absolute value of informativeness (from 0 to 1);
3) The volume of samples in different classes may be different.

The method of cumulative frequencies. This method is applied is cases of classification into two classes. Also the same sample sizes for the two classes are required. The method represents the construction of empirical distributions for objects of both classes in the same coordinate axis. Then for each interval on the coordinate axis the cumulative frequency is calculated (the sum of all frequencies from the first to the current interval). There is a clear analogy with the determination of the probability distribution function by distribution density integration. Assessment of informativeness is the maximum frequency difference for the two classes (among all the intervals).

The method of Kullback. This method was founded by American cryptographer and mathematician Solomon Kulbak. As the informativeness estimation he used the divergence, i.e. the measure of divergence between classes. The peculiarities of this method are the independency from the sample size due to the use of frequencies and the ability to assess the informativeness only for two-class classification. The following formula is used:

$$I(x) = \sum_{i=1}^{G}[P_{i1} - P_{i2}] * \log_2 \frac{P_{i1}}{P_{i2}} \tag{4}$$

$$P_{ik} = \frac{m_{ik}}{\sum_{i=1}^{G} m_{ik}}, k = 1; 2 \tag{5}$$

Where:

G is the number of gradations of the feature *x*;

$m_{i,k}$ is the number of objects of class k, where the feature takes the value of gradation *i*;

$P_{i,k}$ is the proportion of objects of class k, where the feature takes the value of gradation *i*.

V. Informativeness and the statistics collection period

The main goal of this work is to identify the most informative features. It should be noted that a significant impact on the feature informativeness might be brought by the method of statistic collection. In previous work there was the formula of the average frequency of the packet generation in the network: the average frequency is expressed by the algebraic sum of the frequency of packet generating corresponding to different nodes. Reciprocal to the average frequency value is the average period of the new packet generation in the network.

The first experiment was aimed at determination of the dependence between the average period of the new packet generation and the statistics collection period. This has established a model of the network of 15 nodes with mesh topology. The period of the new packet generating by each node is subject to a normal distribution with expected value 10 and standard deviation equal to 1. At the same time to identify possible dependency based on the classification from the packet size, a geometric distribution with parameter 0.8 was introduced. It indicates the number of frames in each packet. On the basis of previous work, the ideal case of the network with an average delay in 0.007168 seconds was considered. The topology of the studied network is shown in figure 1. Each router is PAN of 5 nodes. It is assumed that each PAN has the «star» topology, and transmission is carried out in the channel that is different from the one used for interaction between routers.

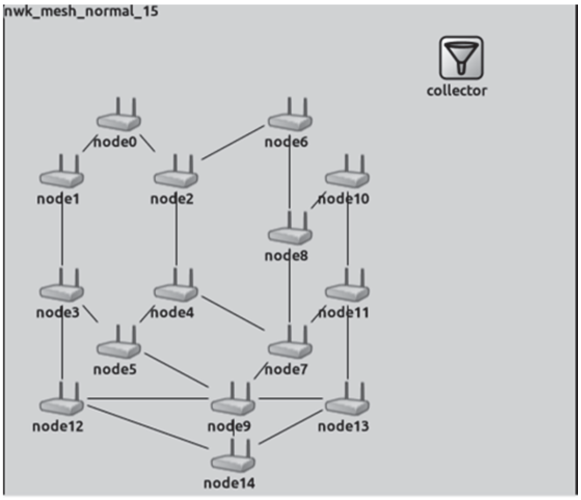

Figure 1. Mesh network topology with 15 routers

During the experiment, the samples with length of 500 records for the following periods for the collection of statistical information (T=10 seconds is the average period of the new packet generating in the network) were obtained:

1) 5 seconds (0.5*T);
2) 10 seconds (1*T);
3) 20 seconds (2*T);
4) 50 seconds (5*T);

5) 100 seconds (10*T);
6) 10 minutes (60*T);
7) 1 hour (360*T).

The results of the experiment are presented in Table IV. For each of the collection periods indicated the most informative feature and its informativeness. Figures 2 and 3 show graphs for the informative periods 10T and 360T.

The analysis of obtained data resulted it the following statements:

1) Informative features significantly depend on the statistics collection period: the longer the period, the greater the value of informativeness;
2) The characteristics that are informative for small periods of collection of statistical information may be uninformative for large periods and vice versa.

TABLE IV. MAXIMUM INFORMATIVENESS AT DIFFERENT PERIODS OF STATISTICS COLLECTION

Period	Feature	Informativeness (by Shannon)
0.5*T	num_packets_equal_src_pan_max	0.0743
1*T	num_forwarded_packets	0.0983
2*T	num_forwarded_packets_max	0.1330
5*T	num_forwarded_packets_max	0.1933
10*T	num_packets_equal_src_pan_min	0.2365
60*T	num_packets_out_avg	0.3594
360*T	num_packets_out_avg	0.4421

Feature screening cannot be done solely on the basis of the method of Shannon: the feature informativeness depends on the ability to divide the sample according to the described classes. The feature that efficiently separates abnormal and normal behaviour may show low informativeness, because it cannot divide objects corresponding to the different ways of attack committing. However, this behaviour may be compensated by the application of compositions of algorithms (i.e., using boosting). Therefore, in addition to evaluation by the method of Shannon for all classes, it is necessary to assess the feature informativeness from the point of view of classification for 2 classes: for all pairs of samples of «normal behaviour»-«abnormal behaviour».

It should be noted that when using features as the basis for the creation of intrusion detection system, it is acceptable to implement the detection of abnormal behaviour for several periods of statistics collection. For example, if the average period of generation of the packet in the system is 10 seconds, it is possible to collect statistics with periods of 10 seconds, 2 minutes, 10 minutes, 30 minutes and 1 hour. For each period its own assessment of the system behaviour is given, and the longer the period is, the more trust the information about the network state returned by the intrusion detection system has.

Theoretically, information may be collected iteratively, i.e. the feature value for the period N*T is formed from the values for N periods of duration T. This approach has its advantages (allows to identify the attacker node after the detection of the attack) but negates the main advantage of intrusion detection system based on statistical methods: the amount of statistic data circulating over the network is too small. A large part of described features can be collected for each PAN separately, and the final value can be calculated on the basis of the received values for the subnets. In this case, only 8 bytes (size of an integer in modern computers) is enough to transfer the major part of the features.

The important fact should be noted regarding the metrics of informativeness derived by using the method of Shannon. In fact, this method is used to assess the ability of the feature to divide the sample into classes. Therefore, if in the sample there is the class that cannot be described with any feature for various reasons, the total value of the informativeness will «sink» for all features. So, in addition it is necessary to assess the ability of each feature to divide the sample only to two classes: normal system operation and system under attack. In order to do this, another experiment was conducted: for each pair of samples of «normal behaviour»-«under attack», values of the informativeness of each feature for three periods – 10T, 60T and 360T - were counted. The results for class of «denial of sleep» attack are presented in Tables V-VII.

TABLE V. THE 5 MOST INFORMATIVE FEATURES ACCORDING TO THE METHOD OF SHANNON

Period	Feature	Informativeness
10T	num_packets_equal_dest_max	0.993084
	num_packets_equal_src_max	0.947495
	num_packets_equal_dest_pan_max	0.326937
	num_packets_out_avg	0.063563
	num_frames_out_avg	0.056685
60T	num_packets_equal_src_max	1
	num_packets_equal_dest_max	1
	num_packets_equal_dest_pan_max	1
	num_packets_created_max	1
	num_packets_out_avg	0.449877
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 VI. THE 5 MOST INFORMATIVE FEATURES ACCORDING TO THE KULLBACK METHOD

Period	Feature	Informativeness
10T	num_packets_equal_dest_max	9.465693477
	num_packets_equal_src_max	6.099388733
	num_packets_equal_dest_pan_max	2.490051765
	num_packets_out_avg	0.547426314
	num_frames_out_avg	0.452548685
60T	num_frames_out_avg	3.578476423
	frac_packets_in_out_avg	2.987830226
	frac_packets_in_out_pan_avg	2.596515492
	num_forwarded_packets_max	2.100943134
	num_packets_out_max	1.919809395
360T	num_frames_avg	8.393924912
	num_packets_equal_dest_pan_min	5.926747972
	num_frames_in_avg	4.500561628
	frac_packets_in_out_avg	3.296978144
	weighted_num_packets_in_max	2.116442256

TABLE VII. THE 5 MOST INFORMATIVE FEATURES ACCORDING TO THE METHOD OF CUMULATIVE FREQUENCIES

Period	Feature	Informativeness
10T	num_packets_equal_dest_max	799
	num_packets_equal_src_max	789
	num_packets_equal_dest_pan_max	505
	num_frames_out_avg	214
	frac_packets_in_out_avg	167
60T	num_packets_equal_src_max	250
	num_packets_equal_dest_max	250
	num_packets_equal_dest_pan_max	250
	num_packets_created_max	249
	num_packets_out_avg	162
360T	num_packets_equal_src_avg	168
	num_packets_equal_src_max	168
	num_packets_equal_dest_max	168
	num_packets_equal_dest_pan_max	168
	num_packets_created_max	168

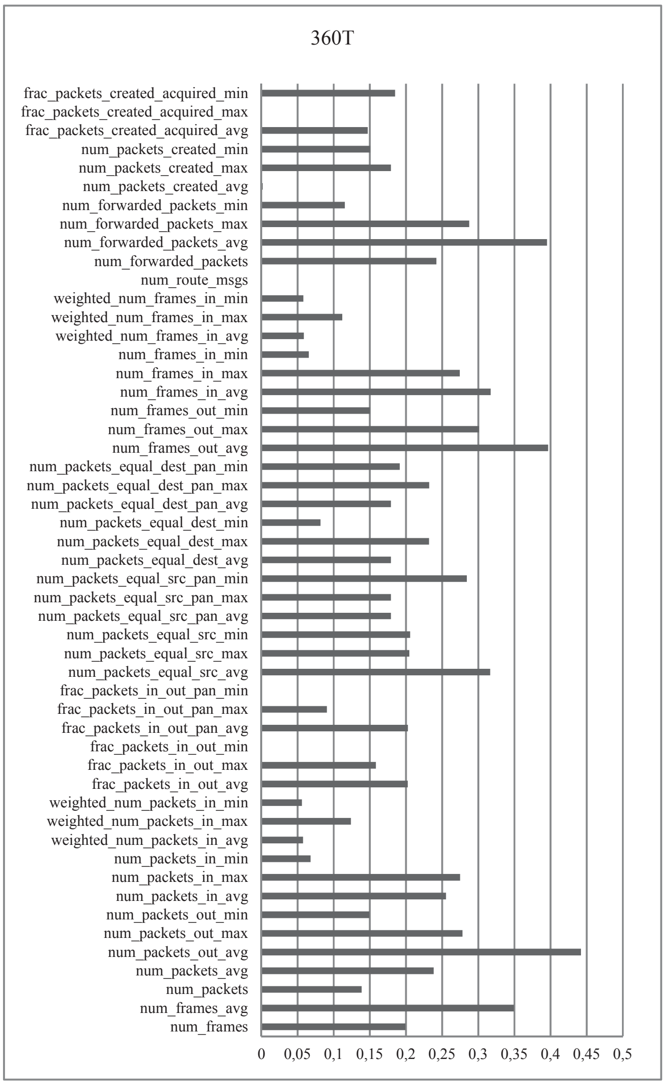

Fig. 2. Feature informativeness with the statistics collection period 360T

Due to the large volume of collected data only the main conclusions are provided:

1) For most attacks (with the exception of variations of retransmissions and selective forwarding), there are characteristics that uniquely distinguish the attack from normal behaviour (it should be noted that this does not guarantee the identification of the specific kind of attack);
2) The longer statistics collection period is, the more the number of maximally informative features is in general case;
3) The methods of Shannon and method of accumulated frequencies most often return the same results, and the results of the method of the Kullback often differ.

Low feature informativeness in case of «re-transmission» and «selective forwarding» attacks can be explained by two factors:

1) For the features-ratio the following assumption is used: if one of the elements of the ratio is 0, then the entire ratio is equal to 0. As a result, for example, for selective forwarding the situation, in which the attacker discards all the packets sent by some node, may occur. Then the number of received packets by the destination node from that node will be 0. The ratio between the number of sent packets by the node and the number of received packet from the node will be equal to 0. That does not differ from the situation when the packets are never passed. The solution is to use the INT_MAX machine constants if the denominator of the ratio takes the value of 0;

2) In order to selective transmission and re-transmission become implemented, it is necessary for packets from the attacked node pass through the attacking node. The experiment was conducted with three cases for each type of attack: either all packets or packets from the particular source or the packets for the particular node were discarded or re-sent. The first case can be tracked even with the period of 60T, but the second and the third. The reason is that the destination address is chosen randomly (from uniform distribution). So, packet, which could be resubmitted or discarded, did not pass through attacking node at all. And the case when the packets pass through the attacking node occurs so infrequently that the change in the characteristic of features-ratio differs a little from the change in the case when transmission to the destination node is not completed at the time of statistics collection. The solution is to study the topology of the network adapted to the attack, when the attacker node receives the most of the packets from the attacked node. Such case occurs often in practice.

VI. THE DEPENDENCE OF THE FEATURE INFORMATIVENESS ON THE SIZES OF THE SAMPLES

The third experiment conducted in the study is to assess the dependence of feature informativeness on the sample size. For example, the methods of Shannon and Kullback operate on frequencies that can change with increase of power of set of objects. The natural consequence is to test the presence of addiction and its formalization.

In order to obtain the answer to the question, three samples of different capacities for the period 2T: 500, 1000 and 2500 (for the normal mode of network operation and each type of attack) were received. Then there was the resulting estimation of the informativeness by the method of Shannon. The results are presented in Table VIII.

Therefore, at high power values of sampling, the informativeness is not increasing with increasing of number of objects. So, the optimal number of learning objects is chosen either empirically, or is dictated by the used method of machine learning. For example, when teaching by the method of stochastic gradient, only a part of the training sample is intentionally used. It is number of objects that are sufficient to achieve the state of the algorithm, in which further gradient steps do not lead to a significant increase in the classification accuracy.

TABLE VIII. THE MOST INFORMATIVE FEATURE WITH DIFFERENT SAMPLE SIZES

The length of the sample (for each attack)	Feature	Informativeness
500	num_forwarded_packets_max	0.1361
	num_packets_out_max	0.1217
	num_forwarded_packets	0.1163
	num_packets_equal_src_pan_max	0.1066
	num_frames_out_max	0.1015
1000	num_forwarded_packets_max	0.1330
	num_packets_out_max	0.1202
	num_forwarded_packets	0.1136
	num_packets_equal_src_pan_max	0.1070
	num_frames_out_max	0.1005
2500	num_forwarded_packets_max	0.1330
	num_packets_out_max	0.1202
	num_forwarded_packets	0.1136
	num_packets_equal_src_pan_max	0.1070
	num_frames_out_max	0.1005

VII. CONCLUSION

In the article the question of formalization of the feature space for the construction of intrusion detection system in wireless sensor networks is studied. Features were allocated on the basis of existing standards and described attacks on wireless sensor networks. Assessment of informativeness was made using three methods: the method of Shannon, the method of Kullback and method of accumulated frequencies. Important conclusions about the dependencies between the informativeness of features, the statistics collection period and sample size are drawn based on the results of the performed experiments. In future, the dependence of the informativeness of the selected features on characteristics of the network, for example, on topology, would be assessed. Subsequently, the most informative features would be used for training machine learning algorithms and for comparing the accuracy of performed classification.

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.