5

Social Network Approach to Anomaly Detection in Network Systems

Grzegorz Kołaczek and Agnieszka Prusiewicz
Wroclaw University of Technology
Poland

1. Introduction

The problem of the network security is taken up since eighties (Denning et al., 1987) and is developed up today (Beltilch et al., 2004, Bera, 2010, Dasgupta, 1999, Basile, 2007, Wilson, 1999). A major problem of automatic intrusion detection is that, it is difficult to make a difference between normal and abnormal user behaviour. Intrusion detection system should not only recognise the previously known patterns of attacks, but also react in case of appearance of the new events that violate the network security policy. The distributed nature of the task of the network security monitoring requires applying of the distributed tools for network security maintaining. The most important postulate addressed to the intrusion detection systems is that, such systems should automatically react in case of detecting the security policy breach to prevent the attack execution or to reduce the potential loss in the network systems. Intrusion detection systems should be equipped with the components responsible for the permanent observation of the states of monitored nodes and components that integrate the results of these observations and diagnose the security level of the system (Kolaczek et al., 2005, Nguyen et al., 2006).

A comprehensive survey of anomaly detection systems is presented in (Patcha & Park, 2007) and a comparison of different approaches to intrusion detection systems is given in (Bejtlich, 2004). One of the first agent systems for network security monitoring has been proposed in works (Balasubramaniyet et al., 1998, Spafford & Zamboni, 2000). In work (Kolaczek et al., 2005) a framework of an original proposal of the intrusion detection system based on the multi-agent approach was presented. In particular, the architecture of such a system and the task of agents were specified. Proposed ideas were further developed and in work (Nguyen et al., 2006) the problem of anomalies detection on the basis of the nodes traffic analysis was discussed. The proposal of the method for Denial of Service Attack detection was given in (Prusiewicz, 2008a).

In this work we propose a novel framework of a multi-agent system for anomaly detection. The originality of our solution consists of applying the social network approach to Man in the Middle Attack (MITM) detection in a network system. Our proposal is based on the social networks discovery and their characteristics measurement to detect anomalies in network traffic. We assume that network communication between nodes constitutes social network of users and their applications, so the appropriate methods of social network formal analysis can be applied. The other important assumption is that values of these social

network parameters for a given node and their distribution for all nodes tend to be constant under normal conditions (Golbeck, 2005, Jamali, 2006, Park, 2007).
We measure the values of the parameters that describe the social network consisted of the nodes and then verify whether the communication patterns between the members of the social network have been violated. The organization of the remaining part of this chapter is as follows. In Section 2 the social networks and the properties of social network are introduced. Then in section 3 the architecture of the multi-agent monitoring system is given. In section 4 the problem of anomaly detection in a social network is taken up. In particular the general schema of anomaly detection procedure is given, the case study of man-in-the-middle attack is carried and the method for this type of attack detection is proposed.

2. Social networks

The basic idea about social networks is very simple. It could be understood as a social structure made of actors which can be represented as network nodes (which are generally individuals or organizations) that are tied by one or more specific types of interdependency, such as values, visions, idea, financial exchange, friends, kinship, dislike, conflict, trade, web links, etc. The resulting structures are often very complex (Butts, 2008, Jamali, 2006, Golbeck , 2005). Social relationships in terms of nodes and ties among them could be used in various types of analysis. A number of academic researches have shown that dependences form social fields play a critical role also in many other fields and could be used in determining the way problems could be solved.

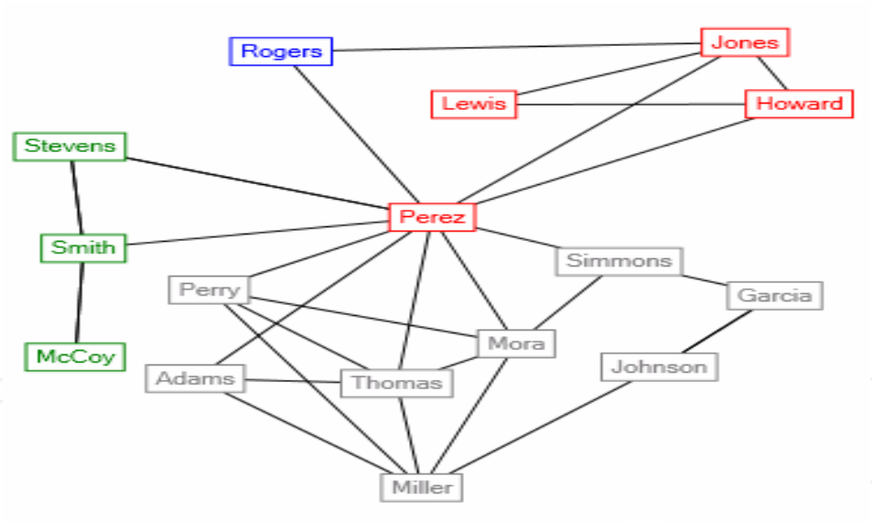


Fig. 1. The example of social network structure (Batchelor, 2010)

Better understanding of social networks requires a complete and rigorous description of a pattern of social relationships as a necessary starting point for analysis. The most convenient situation is when we have complete knowledge about all of the relationships between each pair of actors in the population. To manage all pieces of information related to social network the mathematical and graphical techniques have been used. This formal apparatus allows us to represent the description of networks compactly and systematically. In this context, social network analysts use two kinds of tools from mathematics to represent information about patterns of ties among social actors: graphs and matrices.

Network analysis uses one kind of graphic display that consists of nodes to represent community members and edges to represent ties or relations. There are two general types of situation when there are a single type of relations among the community members and more than one kind of relation. The first one can be represented by the simplex graph while in the second case we use multiplex graphs. Additionally, each social tie or relation represented by graph may be directed or undirected (tie that represents cooccurrence, co-presence, or a bonded-tie between the pair of community members). Another important feature related to the social networks and their graph representation is the strength of ties among community members. In a graph it may be one of the following types: nominal or binary (represents presence or absence of a tie); signed (represents a negative tie, a positive tie, or no tie); ordinal (represents whether the tie is the strongest, next strongest, etc.); or valued (measured on an interval or ratio level).
Other basic social network proprieties that can be formally described and so can constitute a good background for analysis of community dynamics and which can be applied to detect various types of security breaches are as follows (Scott , 2000):

- The Connections between nodes
  The number of immediate connections may be critical in explaining how community members view the world, and how the world views them, so it could be also important factor while modelling trust relations within community. The number and kinds of ties are a basis for similarity or dissimilarity to other community members the direction of connections may be helpful to describe the role of the community member in the society, it can be "a source" of ties, "a sink", or both.

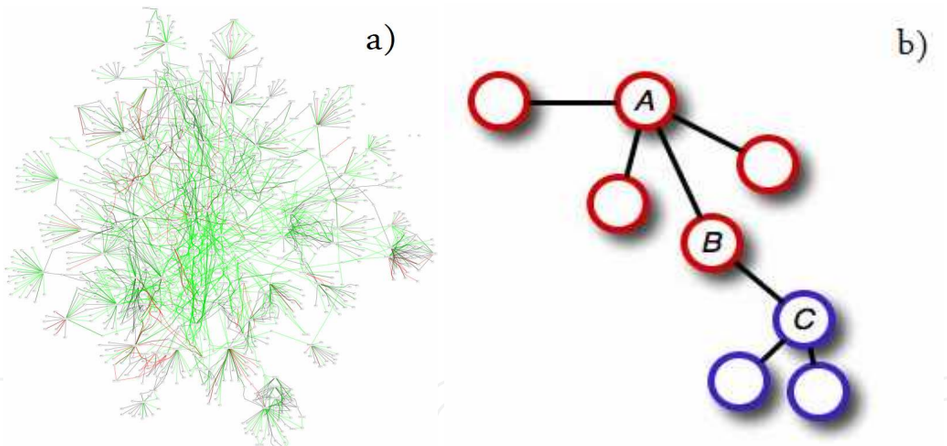


Fig. 2. A network with large number of connection between nodes (a) and small number of connections (b)

- The size of a network
  The size of a network is indexed simply by counting the number of nodes, critical element for the structure of social relations because of the limited resources and capacities that each community member has for building and maintaining ties. The size of a network also influences trust relations while in bigger group it is easier to preserve anonymity and it is more difficult to evaluate trust values.
- The density of a social network
  The density of a social network is defined as the number of existing connections divided by the number of maximum possible connections. The number of logically

possible relationships grows exponentially as the number of actors increases linearly. In communities with greater value of density parameter it should be easier to maintain relations between nodes as we get more information about the other community members.

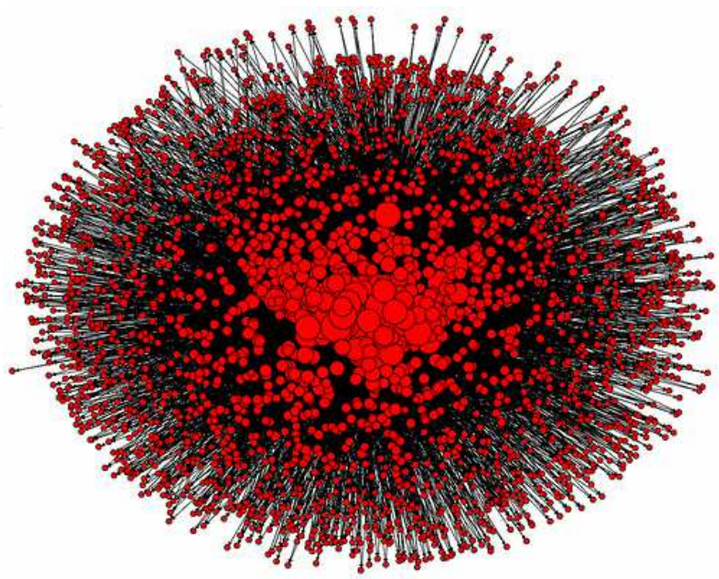


Fig. 3. High density social network (Morrison, 2008)

- The degree of a network node
  It tells us how many connections a community member has. Where out-degree is the sum of the connections from the community member to others and in-degree is the sum of the connections to the particular community member from others. Out-degree and in-degree are also referred as fan-out and fan-in parameters. This type of parameter has been proved to be invariant for a long time periods and different scales (subnet sizes) or traffic types (protocols) of data flows in communication networks (Allmanz, 2005). Experiments showed that both Fan-in and Fan-out for a given node and their distribution for all nodes tend to be constant under normal conditions. While network is affected by some type of attack the structure of communication is often heavily affected and the distribution changes. There is also a detectible dependence between type of the attack and communication pattern disturbance (Kohler, 2002). At the other hand, community members that receive information from many sources may also be more powerful community members. However, these nodes could also suffer from "information overload" or "noise and interference" due to contradictory messages from different sources. Impact for the social relations between nodes is similar to that described in a case of density, dependently from in/out-degree when an individual has more or less information about its neighbourhood.
- The reachability of community members
  A community member is "reachable" by another if there exists any set of connections by which we can find link from the source to the target entity, regardless of how many others fall between them. If some community members in a network cannot reach others, there is the potential of a division of the network. For example, disconnected community members could have more problems to evaluate trust value.

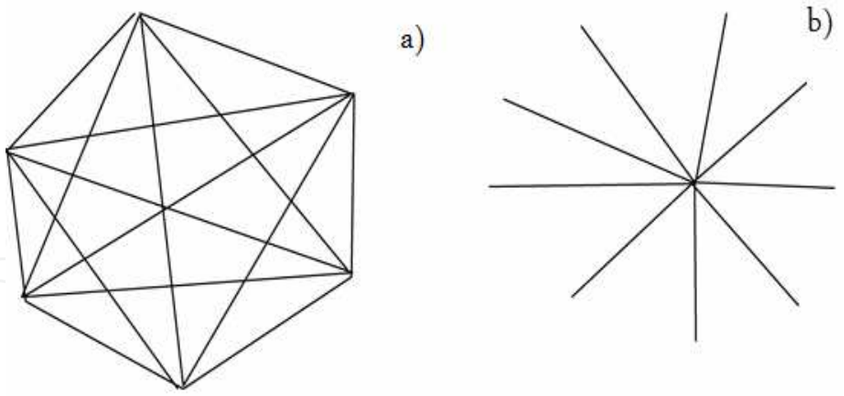


Fig. 4. The examples of networks with different node degree parameter values (a) Perfect graph of n nodes with avg. node degree n-1. (b) Star graph of n+1 nodes with avg. node degree (2n)/(n+1)

- The transivity of network nodes connections
  The transitivity principle holds that, if A is tied to B, and B is tied to C, then A should be tied to C. The triadic relationships (where there are ties among the actors) should tend toward transitivity as an equilibrium condition. One of the most important type of social relation as trust is not strictly transitive and so this propriety not necessarily influences trust evaluation process.
- The distance between the network nodes
  An aspect of how individuals are embedded in networks, two actors are adjacent when the distance between them is one. How many community members are at various distances from each other can be important for understanding the differences among community members in the constraints and opportunities they have as a result of their network location. Community members located more far apart from each other in the community have more problems with establishing new relations than the members, which are close.
- The geodesic distance between the network nodes
  The geodesic distance is defined as the number of relations in the shortest possible walk from one community member to another. Many algorithms in network analysis assume that community members will use the geodesic path when communicating with each other.
- The diameter of a network
  The diameter of a network is the largest geodesic distance in the connected network which tells us how "big" the community is, in one sense quantity in that it can be used to set an upper bound on the lengths of connections that we study.
- The cohesion of a social network
  The degree to which nodes are connected directly to each other by communication links. Count the total connections between actors more strong connection between community members should determine grater trust values.
- Centrality, Power, Betweenness
  Centrality, closeness, betweenness describe the locations of individuals in terms of how close they are to the "center" of the action in a network – though there are a few different definition of what it means to be at the canter. The more important community member is, the more important its opinions should be.

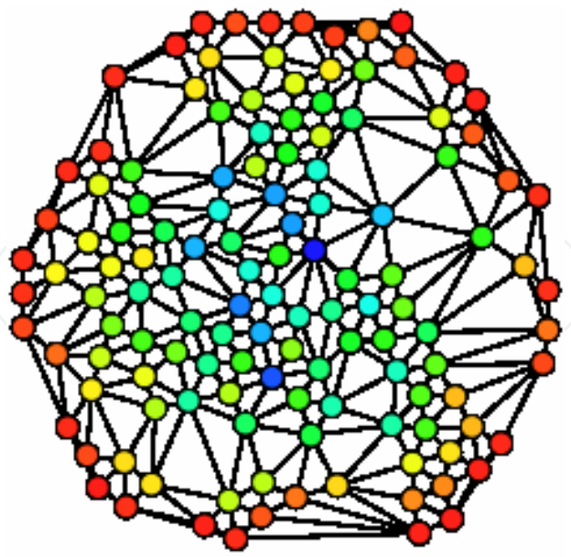


Fig. 5. Example of betweenness centrality. Red colour indicates the lowest betweenness centrality value and blue the highest (Bailin, 2009)

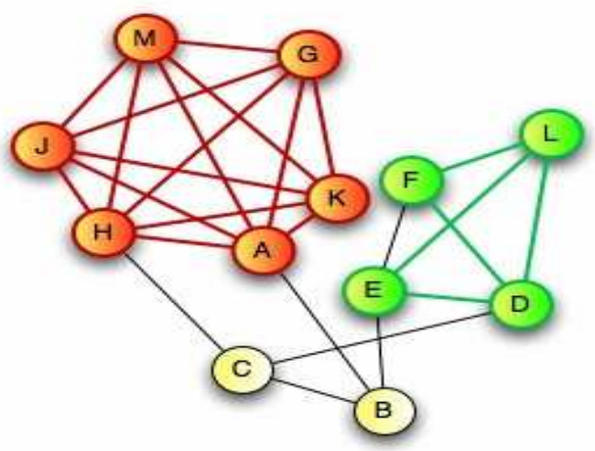


Fig. 6. Cliques example. Network with two cliques – first one is composed of nodes: A,G,H,J,K,M, the second: D,E,F,L

- The eigenvector of the geodesic distances
  An effort to find the most central community members in terms of the "global" or "overall" structure of the network, and to pay less attention to patterns that are more "local"
- The cliques in the network
  A subset of community members who are more closely tied to each other. A clique typically is a subset of community in which every node is connected to every other node of the group and which is not part of any other clique. Idea of cliques within a network is a powerful tool for understanding social structure and the embeddedness of individuals. Cliques reflect the groups of community members with strong relationship.

So, sudden change in communication pattern within such a group may be related to security breaches.

- The clustering coefficient
  The probability that two nearest neighbours of a given node are also neighbours of each other. The value of clustering coefficient provides a quantitative measure for cliques in communication graph.

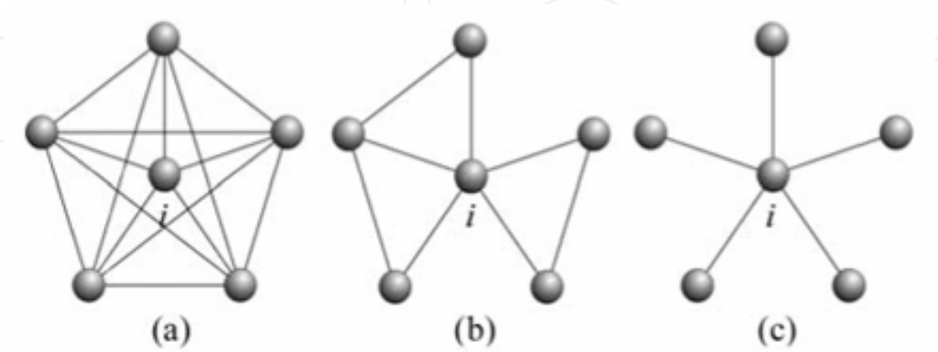


Fig. 7. Example of clustering coefficients (cc) for different networks. a) cc=1, b) cc=0.3, c) cc=0

3. The architecture of the multi-agent monitoring system

It is assumed that there are two layers in the architecture of the multi-agent monitoring system: monitoring layer and control layer (Fig. 8). Monitoring layer consists of the nodes that are monitored by the monitoring agents. While control layer consists of the security control agents that are responsible for collecting data from monitoring agents and determining general characteristics of the network traffic in the monitored region. These characteristics describe communication patterns in the monitored region. We assume that communicating nodes constitutes the social network. Each security control agent is responsible for controlling one region (social network).
The patterns of discovered social networks are temporally collated by security control agents with communication patterns stored in security control agents private databases in order to verify if any security policy breach has been occurred.

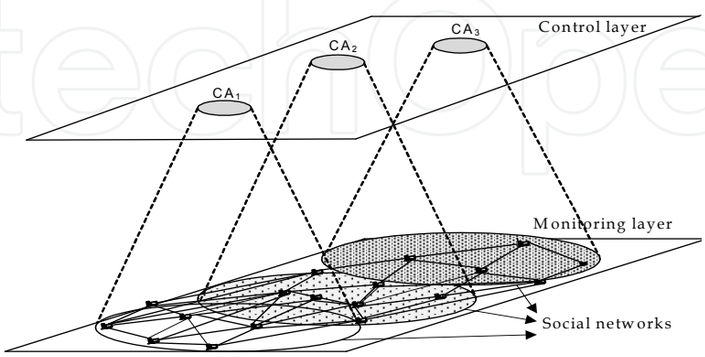


Fig. 8. Two-layers multi-agent monitoring system architecture

Before the internal organization of the monitoring and security control agents will be given, let us denote: V as the set of the nodes, $V = \{v_1, v_2, \ldots, v_k, \ldots, v_K\}$, $K \in N$, MA

($MA = \{MA^1, MA^2, \ldots, MA^k, \ldots MA^K\}$) as the set of the monitoring agents, SA ($SA = \{SA^1, SA^2, \ldots, SA^g, \ldots SA^G\}$), $G \in N$ as the set of security control agents, SN ($SN = \{SN^g : SN^g \subseteq V\}$) as the set of the social networks (monitoring regions) and $P = \{P_1, P_2, \ldots, P_Z\}$ as the set of the observed parameters describing the nodes from V.

3.1 Monitoring agent's internal organization

Each monitoring agent $MA^k \in MA$ observes the states of one node from V in their monitoring regions (social networks) from SN with the reference to the values of the parameters from the set P. The results of the observations are captured in their private set of observations.

Definition 1. A single observation of agent MA^k is stored as a tuple [Prusiewicz, 2008a]:

$$O^k\left(\left(P_j, x\right), t_n\right) \in DB^k \tag{1}$$

where: $P_j \in P$, $t_n \in T$ and T is the universe of the timestamps and DB^k denotes the database of the agent MA^k.

Such observation refers to the situation that at the timestamp t_n the agent MA^k has observed in the node v_k the value of the parameter P_j equals x.

3.2 Security control agents internal organization

Security control agents control the monitoring regions. The size of the monitoring regions may change by adding or removing nodes as a consequence of the social networks evolutions. Security control agent SA^g, $SA^g \in SA$ is built from three modules: Data Storage Module, Social Network Module and Security Control Module. Security control agent SA^g collects data from databases of the monitoring agents and builds communication matrix in Data Storage Module [Prusiewicz, 2008b].

Definition 2. The communication matrix CM^g is defined as:

$$CM^g = \left[a_{mn}\right]_{GxG} \tag{2}$$

where a_{mn} is the set of time stamps of communication acts between nodes v_m and v_n. The node v_m is a sender and v_n - receiver.

	v_1	v_2	v_5	v_{12}	v_{13}	v_{17}	v_{31}
v_1							
v_2					a_{213}	a_{217}	
v_5			$(t_2, t_5, t_9, t_{23}, t_{28}, t_{34})$				
v_{12}							
v_{13}							
v_{17}			a_{175}				
v_{31}							

Table 1. En example of communication matrix: the node v_2 communicated with the node v_{12} at the timestamps: t_2, t_5, t_9, t_{23}, t_{28}, t_{34}

On the basis of data from communication matrix CM^g the values of the parameters describing the social network SN^g, $SN^g \in SN$ are determined in Social Network Module. Additionally in Social Network Module the patterns of communication between nodes are determined that a are the basis for the social networks discovery. In Security Control Module the procedures for anomalies detections are implemented. In this case the procedure for Man-In-The Middle attack is implemented.

3.3 Determining of the social network characteristics

Social network characteristics are determined on the basis of the data from communication matrix CM^g by the Social Network Module. In this module two data structure are used in order to control the security of the monitoring region: Social Network Patterns and Temporal Communication Patterns defined as follows:

Definition 3. A Social Network Patterns is defined as:

$$SNP^g_{[t_b,t_e]} = \left\langle f^{[t_b,t_e]}_{in,v_i}, f^{[t_b,t_e]}_{out,v_i}, cl^{[t_b,t_e]}_{v_i}, c^{[t_b,t_e]}_{v_i} \right\rangle \tag{3}$$

where:

- $f^{[t_b,t_e]}_{in,v_i}$ is the number of nodes that originate data exchange with node v_i during observation period $[t_b, t_e]$
- $f^{[t_b,t_e]}_{out,v_i}$ is the number of nodes to which v_i initiates conversations during observation period $[t_b, t_e]$
- $cl^{[t_b,t_e]}_{v_i}$ is the clustering coefficient defined according to the following equation:

$$cl^{[t_b,t_e]}_{v_i} = \frac{2\left|E\left(G1\left(v_i^{[t_b,t_e]}\right)\right)\right|}{\deg\left(v_i^{[t_b,t_e]}\right)\left(\deg\left(v_i^{[t_b,t_e]}\right)-1\right)} \tag{4}$$

where:

- $f^{[t_b,t_e]}_{in,v_i}$ is the number of nodes that originate data exchange with node v_i during observation period $[t_b, t_e]$
- $\deg(v_i^{[t_b,t_e]})$ – denotes degree of node v_i during observation period $[t_b, t_e]$
- $G1(v_i^{[t_b,t_e]})$ – is the set of nodes which are connected with v_i via single link (its immediate neighbors) during observation period $[t_b, t_e]$
- $E(G1(v_i^{[t_b,t_e]}))|$ – is the number of edges among nodes in 1–neighbourhood of node v_i during observation period $[t_b, t_e]$
- $c^{[t_b,t_e]}_{v_i}$ is the centrality of the node, it describes the temporal location of the node v_i during observation period $[t_b, t_e]$ in terms of how close it is to the "canter" of the action in a network.

There are four measures of centrality that are widely used in network analysis: degree centrality, betweenness, closeness, and eigenvector centrality. The proposed method uses Eigenvector centrality measure which assigns relative scores to all nodes in the network based on the principle that connections to high-scoring nodes contribute more to the score of

the node in question than equal connections to low-scoring nodes. For examle Google's PageRank is a variant of the Eigenvector centrality measure (Page, 1998). For the node v_i the centrality score is proportional to the sum of the scores of all nodes which are connected to it within observation period $[t_b, t_e]$:

$$c_{v_i}^{[t_b,t_e]} = \frac{1}{\lambda} \sum_{j \in M^{[t_b,t_e]}(i)} v_j^{[t_b,t_e]} = \frac{1}{\lambda} \sum_{j=1}^{N} A_{i,j}^{[t_b,t_e]} v_j^{[t_b,t_e]} \tag{5}$$

where:

- $M^{[t_b,t_e]}(i)$ is the set of nodes that are connected to the node v_i during observation period $[t_b, t_e]$,
- N is the total number of nodes,
- $A_{i,j}^{[t_b,t_e]}$ is the adjacency matrix of the network during observation period $[t_b, t_e]$,
- λ is a constant.

Definition 4. A Temporal Communication Patterns is the set of social network characteristics that has been determined at the time intervals, defined as:

$$TCP^g = \left\langle TCP^g_{[t_b',t_e']} \middle| TCP^g_{[t_b',t_e']} = \left\langle f_{in,v_i}^{[t_b',t_e']}, f_{out,v_i}^{[t_b',t_e']}, cl_{v_i}^{[t_b',t_e']}, c_{v_i}^{[t_b',t_e']} \right\rangle \right\rangle \tag{6}$$

where each element of the set TCP^g has the same structure as Social Network Patterns. The difference is that the values of $SNP^g_{[t_b,t_e]}$ are the patterns that describe the monitored social network SN^g ($SN^g \subseteq V$). They are discovered on the basis of the historical network traffic data analysis. Social Network Patterns are used to discover any security policy breaches in a network system. While the values from Temporal Communication Patterns describe the current communication characteristics of monitored region. The last element of the TCP^g is a current characteristics of communication patterns of a social network SN^g.

Having the values of the parameters from Social Network Patterns and current social network characteristics the procedure for anomaly detection might be applied.

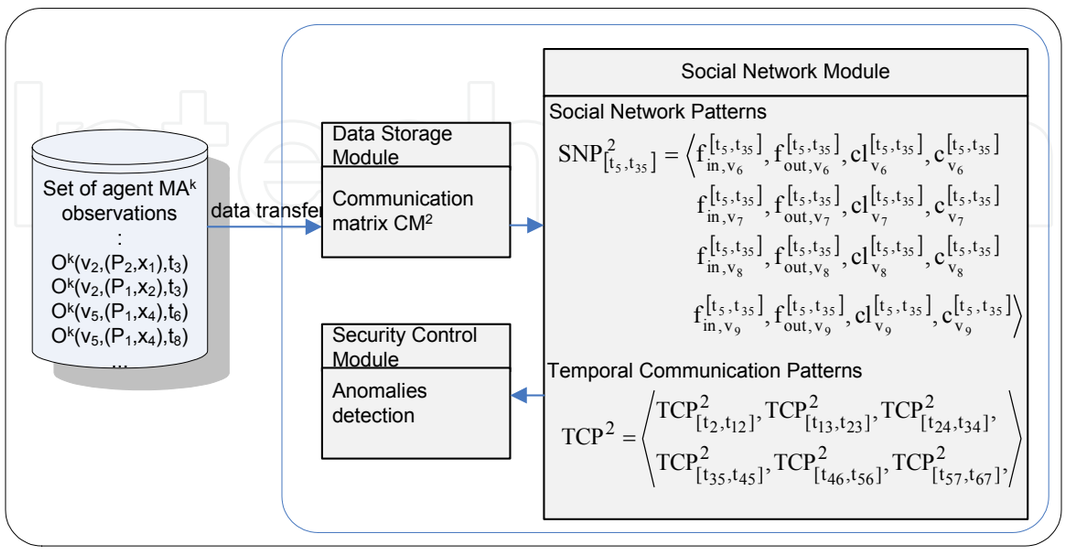


Fig. 9. The process of determining of the social network characteristics and anomaly detection

On the fig. 9 the process of anomaly detection carried out by the security control agent is illustrated. First the observations from the monitoring agents embodied in the nodes v_6, v_7, v_8, v_9 are captured by the Data Storage Module and used to determine communication matrix CM^2. Data from CM^2 are sent to Social Network Module, responsible for determining the patterns of communications in an observed network. The social network patterns $SNP^2_{[t_5,t_{35}]}$ have been determined for the nodes: v_6, v_7, v_8, v_9 and the time interval $[t_5, t_{35}]$ to control the security. The current communication patterns TCP^2 are compared with $SNP^2_{[t_5,t_{35}]}$ to control the security of SN^2.

4. Man-in-the-middle attack detection

The man-in-the-middle attack (often abbreviated MITM) is a form of active eavesdropping in which the attacker makes independent connections with the victims and relays messages between them, making them believe that they are talking directly to each other over a private connection when in fact the entire conversation is controlled by the attacker. To perform the effective attack, the attacker must be able to intercept all messages going between the two victims and inject new ones, which is straightforward in many circumstances (Fields, 1995).

This type of attack can be as analyzed as a general problem resulting from the presence of intermediate parties acting as proxy for clients on either side (Asokan, 2002, Shim, 2003, Welch, 2003). The problems related to the MITM attacks are also related to trust relation among geographically distributed subjects. If communicating with each other subjects are trustworthy and competent the risk of the MITM attack is low. If communicating parts do not know each other or has no trust relation, the risk of the attack increases. By acting as proxy and appearing as the trusted client to each side, the intermediate attacker can carry out much mischief, including various attacks against the confidentiality or integrity of the data passing through it. So, one of the most urgent question is how one can detect MITM attacks.

It is important to notice, that MITM attack is a general security problem not only related to cryptographic applications. An example of such non-cryptographic man-in-the-middle attack was caused by one version of a Belkin wireless network router in 2003 (Leyden, 2003). This router periodically would take over an HTTP connection being routed through it: it would fail to pass the traffic on to destination, but instead itself respond as the intended server. The reply it sent, in place of the requested web page, was an advertisement for another Belkin product. This 'feature' was removed from later versions of the router's firmware (Scott, 2000).

Another example of such type of man-in-the-middle attack could be the "Turing porn farm". This schema of the attack potentially could be used by spammers to defeat CAPTCHAs (Petmail). The general idea is that the spammer sets up a pornographic web site where access requires that the user solves the CAPTCHAs in question. However, this attack is merely theoretical because there is no evidence of building Turing porn farm by the time being (Atwood, 2006). There are available several ready to use tools which implement the MITM idea and which can be used for communication interception in various environments, e.g. dsniff – a tool for SSH and SSL MITM attacks, Ettercap - a tool for LAN based MITM attacks, AirJack - a tool that demonstrates 802.11 based MITM attacks, and many others.

4.1 Evaluation of MITM event probability value

We assume tracking four communication patterns: Fan-in (from here on denoted as $f_{in,v_i}^{\Delta t}$ for node v_i during the observation period Δt), Fan-out ($f_{out,v_i}^{\Delta t}$), clustering coefficient ($cl_{v_i}^{\Delta t}$) and centrality ($c_{v_i}^{\Delta t}$).

According to the assumption presented by (Allmanz, 2005) that these types of parameter has been proved to be invariant for a long time periods and different subnet sizes or traffic types of data flows, the risk of the MITM incident will be estimated as the abnormal change of the characteristic parameters values for a given social network member.

Let us assume that the collected history record consists of a number of observations of Fan-in values from some starting point up to current time t. So we have $f_{in,v_i}^{\Delta t_1}, f_{in,v_i}^{\Delta t_2}, f_{in,v_i}^{\Delta t_3}, \ldots, f_{in,v_i}^{\Delta t_k}$. Now, consider the Fan-in as a random variable F_{in,v_i}. Thus, $(f_{in,v_i}^{\Delta t_1}, f_{in,v_i}^{\Delta t_2}, f_{in,v_i}^{\Delta t_3}, \ldots, f_{in,v_i}^{\Delta t_k})$ is a sample of size k of F_{in,v_i}. We also assume all of the $f_{in,v_i}^{\Delta t}$ to be independent. It is commonly known that the mean value and the variance of F_{in,v_i} can be estimated by using the following formulae:

$$\overline{F}_{in,v_i} = \frac{1}{m}\sum_{j=1}^{k} f_{in,v_i}^{\Delta t_j} \tag{7}$$

$$S_{in,v_i} = \frac{1}{k-1}\sum_{j=1}^{k} (f_{in,v_i}^{\Delta t_j} - \overline{F}_{in,v_i})^2 \tag{8}$$

$\overline{F}_{in,v_i}$ and S_{in,v_i} are thus the estimations (based on the data being at our disposal) of mean value and the variance of F_{in,v_i}. Obviously the bigger our sample is, the better they approximate $E(F_{in,v_i})$ (expected value of random variable) F_{in,v_i} and $Var(F_{in,v_i})$ (variance of random variable F_{in,v_i}) respectively. From this point we assume that the observations 'number is big enough to state that $E(F_{in,v_i})$ and $Var(F_{in,v_i})$ are known.

Let also $E(F_{out,v_i})$ and $Var(F_{out,v_i})$ for the Fan-out, as well as $E(cl_{v_i}^{\Delta t})$ and $Var(cl_{v_i}^{\Delta t})$ for clustering coefficient and centrality $E(c_{v_i}^{\Delta t})$, $Var(c_{v_i}^{\Delta t})$ be defined in the same way.

In our approach, we will detect the possible MITM events by evaluation of some weighted value related to mean value and variance of fan-in, fan-out, clustering coefficient and centrality. At this stage of research we assume that we will analyze all four parameters independently. This means that it is enough to assume MITM incident if only one of the parameters exceeds threshold value.

From the Chebyshev's inequality we can estimate the upper bound of the probability that $|\overline{F} - x|$ is greater than kS. Where $\overline{F}$ and S are mean value and the variance of X, while X denotes the random variable related to x (in this case one of the following: $f_{in,v_i}^{\Delta t}, f_{out,v_i}^{\Delta t}, c_{v_i}^{\Delta t}, cl_{v_i}^{\Delta t}$).

According to this estimation the probability expectation $E(\varpi_{v_i})$ value of the MITM event for a given parameter will be evaluated using the following formula:

$$E(\omega_{v_i}) = 1 - \frac{1}{\alpha k^2} \tag{9}$$

Where α is a coefficient, which value should be set during a process of tuning-up the detection system to the real network conditions. Parameter k is defined as follows:

$$k = \begin{cases} 1 & \text{if} \left\lfloor \frac{\bar{F} - x}{\sqrt{S}} \right\rfloor < 1 \\ \left\lfloor \frac{\bar{F} - x}{\sqrt{S}} \right\rfloor & \text{if} \left\lfloor \frac{\bar{F} - x}{\sqrt{S}} \right\rfloor \geq 1 \end{cases} \tag{10}$$

4.2 The procedure of the Man-in-the-middle attack detection

Our approach of MITM attack detection has been dedicated especially to effectively detect automated attacks of this type. For example this method should be convenient for detection HoneyBot-based attacks as it has been described in their work by researchers from Institut EURECOM in France (Lauinger, 2010), who are working on automation of social engineering attacks on social networks.

French researchers have developed an automated social engineering tool that uses a man-in-the middle attack and strikes up online conversations with potential victims. In the work (Lauinger, 2010) the proof-of-concept HoneyBot has been presented that poses convincingly as a real human in Internet Relay Chats (IRC) and instant messaging sessions. It lets an attacker collect personal and other valuable information from victims via these chats, or tempt them into clicking on malicious links. The researchers had proved the feasibility and effectiveness of their MITM attack variant. During the tests they were able to get users to click onto malicious links sent via their chat messages 76 percent of the time.

We propose the following idea of algorithm for MITM detection using social network patterns.

Input: D – set of data that can be used to derive and observe patterns of the social network (e.g. e-mail logs, chat rooms records, network traffic, etc.)

Output: $R \in \{Y,N\}$ – information about the social network state according to risk of MITM incidents

BEGIN

1. Take the data set D and derive the social network structure (e.g. using one of the approach presented in section 3.1)
2. For each node find the current value of the monitored social network patterns (fan-in, fan-out, centrality, clustering)
3. Analyze the history of network patterns changes. As we treat the network patterns values as the realization of the random variable, mean value and variance will be calculated for each pattern.
4. For each node compare the latest change of the patterns values to the assumed threshold value.
5. If the result of the step 4 is that the observed parameter value exceeded the threshold, the value of the result variable is set to Y – the high risk of the MITM incident, otherwise it is set to N – small risk of MITM incident.
6. Return to step 2.

END

Remarks:

- due to the social network dynamics, we may consider some periodic more thorough updates of the network structure; it could be represented in the above algorithm by

adding a time related condition in step 6 and then by return to step 1 instead of returning to step 2

- it is possible to consider situation when we are interested only in monitoring for one particular or some specific subset of all nodes (bank client, chat room participant, etc.), then we may investigate some additional information about its activity and use some data fusion methods to improve the accurateness of the final decision (e.g. we may concurrently track the node's activity within several different social network and so build more comprehensive profile of the network identity).
- we should consider if some "suspicious" behaviour in the context of the only one observed parameter is enough to assume MITM incident or we would prefer to wait for more premises or else we will combine the values of all parameters and only after using data fusion methods set up the final decision.

The algorithm for MITM attack detection

Input: Social Network Patterns $SNP^{g}_{[t_b,t_e]}$

Temporal Network Patterns $TNP^{g}_{[t_{b'},t_{e'}]}$

Threshold values: F_{in,v_i-max}, F_{out,v_i-max}, C_{v_i-max}, Cl_{v_i-max}.

Output: The risk of the MITM incidents in the nodes of SN^{g}

BEGIN

1. For each node $v_i \in SN^{g}$ determine the probability expectation values: $E(\omega_{F_{in,v_i}})$, $E(\omega_{F_{out,v_i}})$, $E(\omega_{Cl_{v_i}})$, $E(\omega_{C_{v_i}})$ according to the formula 9.
2. If $E(\omega_{F_{in,v_i}}) > F_{in,v_i\text{-}max}$ or $E(\omega_{F_{out,v_i}}) > F_{out,v_i\text{-}max}$ or $E(\omega_{Cl_{v_i}}) > Cl_{v_i\text{-}max}$ or $E(\omega_{C_{v_i}}) > c_{v_i\text{-}max}$

   then the risk of MITM incident in the node V_i: R_{v_i} := {Y} else R_{v_i} := {N}

END

5. Conclusion

Generally the aim of the network security systems is to protect computational and communication resources from any security policy breaches. Such systems should be equipped with the meachnisms for permanent monitoring the values of the parameters describing their states in order to diangose and protect of their resources. The most important postulate addressed to the intrusion detection systems is that, such systems should automatically react in case of detecting the security policy breaches to prevent the attack executions or to reduce the potential loss in the network systems. Although the problem of the network security has been studied for decades and several methods and approaches have been proposed there is still open problem how to differentiate normal and abnormal states of the network system. In this work we proposed the social network approach to evaluate the security state of the network. The values of the chosen coefficients that characterise the users behaviour are used to discover security breaches occurrence. The idea of our proposal is as follows. First the user behaviours are monitored and the social networks are discovered. Then having the pattern values of social networks characteristics we are able to compare them with the current observations and detect any aberrances.

We proposed two-layers multi-agent system for security monitoring and the algorithm for MITM attack detection.

6. Acknowledgements

The research presented in this work has been partially supported by the European Union within the European Regional Development Fund program no. POIG.01.03.01-00-008/08

7. References

Allmanz M. et.al. (2005). A First Look at Modern Enterprise Traffic, In *Proc. Internet Measurement Conference*, pp. 217-231.

Asokan N, Niemi V, Nyberg K (2002) Man-in-the-middle in tunnelled authentication protocols, *Technical Report 2002*/163, IACR ePrint archive

Atwood J., (2006). CAPTCHA Effectiveness, http://www.codinghorror.com/, last access 01 September 2010

Bailin A., (2009). Measuring digital engagement, http://coi.gov.uk/blogs/digigov/, last access 01 September 2010

Balasubramaniyan, J.S., Garcia-Fernandez, J.O., Isacoff, D., Spafford, E., Zamboni, D. (1998). An Architecture for Intrusion Detection Using Autonomous Agents, Proceedings of the 14th Annual Computer Security Applications Conference

Basile C., Lioy A., Prez G. M., Clemente F. J. G., and Skarmeta A. F. G. (2007). POSITIF: a policy-based security management system, In 8th *IEEE International Workshop on Policies for Distributed Systems and Networks* (POLICY07), pp.280-280

Batchelor J., (2010). What the naked eye cannot see, http://thesocionomicaffect.wordpress.com/, last access 01 September 2010

Bejtlich, R. (2004). Tao of Network Security Monitoring, The: Beyond Intrusion Detection, Addison-Wesley.

Bera P., Ghosh SK., Dasgupta P. (2010) A Spatio-Temporal Role-Based Access Control Model for Wireless LAN Security Policy Management, *4th International Conference on Information Systems,Technology and Management* (ICISTM 2010), LNCS Springer Berlin, vol.54, pp.76-88

Biermann, E., Cloete, E., Venter, L.M., (2001). A comparison of Intrusion Detection systems, *Computers and Security*, vol. 20 Issue: 8, pp. 676-683

Butts J. and Carter T. (2008). Social network analysis: A methodological introduction. Asian *Journal of Social Psychology* 11 (1), 13-41.

Dasgupta, D. (1999). Immunity-Based Intrusion Detection. System: A General Framework. *Proceedings of the 22nd National Information Systems Security Conference*, USA

Denning, D.E, Edwards, D.L, Jagannathan, R., Lunt, T.F., Neumann, P.G. (1987). A prototype IDES: A real-time intrusiondetection expert system. Technical report, Computer Science Laboratory, SRI International, Menlo Park.

Fields A., Ringel M. (1995). Who do you trust? Trust Hierarchies Over Insecure Communications Channels

Golbeck J. and Hendler J. (2005) Inferring trust relationships in web-based social networks, *ACM Transactions on Internet Technology*, pp. 145-165

Jamali M., Abolhassani H. (2006). Different Aspects of Social Network Analysis, 2006. [Online]. Available: http://dx.doi.org/10.1109/WI.2006.61

Kohler E., Liy J., Paxson V., Shenker S. (2002). Observed Structure of Addresses in IP Traffic, In *Proc. SIGCOMM Internet Measurement Workshop*, pp. 253 - 266.

Kołaczek, G., Prusiewicz, A., Juszczyszyn, K., Grzech, A., Katarzyniak, R., Nguyen, N.T. (2005). A mobile agent approach to intrusion detection in network systems, Lecture Notes in Computer Science, *Lecture Notes in Artificial Intelligence*. vol. 3682, pp. 514-519

Lauinger T., Pankakoski F., Balzarotti D., Kirda E. (2010). Honeybot: Your Man in the Middle for Automated Social Engineering, *3rd USENIX Workshop on Large-Scale Exploits and Emergent Threats* (LEET), San Jose

Leyden J. (2003). Help! my Belkin router is spamming me, http://www.theregister.co.uk, last access 01 September 2010

Morrison M.(2008). Map of top 50 UK PR twitter people and their followers http://blog.magicbeanlab.com/, last access 01 September 2010

Nguyen, N.T., Juszczyszyn, K., Kołaczek, G., Grzech, A., Prusiewicz, A., Katarzyniak, R. (2006). Agent-based approach for distributed intrusion detection system design, *Lecture Notes in Computer Science*, vol. 3993, pp. 224-231

Onnela J.P., Saramaki, J., Szabo, G., Lazer, D., Kaski, K., Kertesz, J., Barabasi, Hyvönen, A.L. (2007). Structure and tie strengths in mobile communication networks, *Proceedings of the National Academy of Sciences* 18, 7332-7336, 2007.

Page L., Brin S., Motwani R., Winograd T. (1998). The PageRank Citation Ranking: Bringing order to the Web. Technical Report, Computer Science Department, Stanford University (1998).

Park J. and A. L. Barabási (2007). Distribution of node characteristics in complex networks. *Proceedings of the National Academy of Sciences of the United States of America* 104 (46), 17916-17920.

Patcha, A., Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends, *Computer Networks*, vol. 51, Issue: 12, pp. 3448-3470

Petmail Documentation: Steal People's Time To Solve CAPTCHA Challenges, http://petmail.lothar.com/, last access 01 September 2010

Prusiewicz, A. (2008a). On some method for intrusion detection used by the multi-agent monitoring system, *Lecture Notes in Computer Science*, vol. 5103, pp. 614-623, Kraków, Poland

Prusiewicz, A. (2008b). A multi-agent system for computer network security monitoring. *Lecture Notes in Artificial Intelligence*. 2008, vol. 4953, s. 842-849.

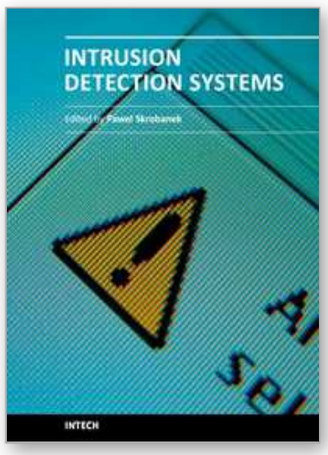
Scott J. (2000). Social Network Analysis: A Handbook. 2nd Ed. Sage, London

Shim K, (2003) A man-in-the-middle attack on Nalla-Reddy's ID-based tripartite authenticated key agreement protocol, http://eprint.iacr.org/2003/115.pdf, last access 01 September 2010

Spafford, E., Zamboni, D. (2000). Intrusion detection using autonomous agents, Computer Networks. *The International Journal of Computer and Telecommunications Networking*, vol.34 , Issue 4, pp. 547-570

Welch, D., Lathrop, S. (2003). Wireless security threat taxonomy, Information Assurance Workshop, *IEEE Systems, Man and Cybernetics Society* , vol., no., pp. 76- 83

Wilson, E. (1999). Network Monitoring and Analysis: A Protocol Approach to Troubleshooting, Prentice Hall.



Intrusion Detection Systems
Edited by Dr. Pawel Skrobanek

ISBN 978-953-307-167-1
Hard cover, 324 pages
Publisher InTech
Published online 22, March, 2011
Published in print edition March, 2011

The current structure of the chapters reflects the key aspects discussed in the papers but the papers themselves contain more additional interesting information: examples of a practical application and results obtained for existing networks as well as results of experiments confirming efficacy of a synergistic analysis of anomaly detection and signature detection, and application of interesting solutions, such as an analysis of the anomalies of user behaviors and many others.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Grzegorz Kołaczek and Agnieszka Prusiewicz (2011). Social Network Approach to Anomaly Detection in Network Systems, Intrusion Detection Systems, Dr. Pawel Skrobanek (Ed.), ISBN: 978-953-307-167-1, InTech, Available from: http://www.intechopen.com/books/intrusion-detection-systems/social-network-approach-to-anomaly-detection-in-network-systems

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen