

QUANTITATIVE FRAMEWORK OF RANDOMLY ROVING AGENTS

H. L. ASLANYAN*

Chair of Discrete Mathematics and Theoretical Informatics, YSU

Quantitative characterization of randomly roving agents in wireless sensor networks (WSN) is studied. Above the formula simplifications, regarding the known results, it is shown that the basic agent model is stochastically equivalent to a similar simpler model. Then a formula for frequencies is achieved in terms of combinatorial second kind Stirling numbers. At allows to justify the roving agents quantitative characteristics.

Keywords: wireless, sensor, network, roving agent.

Introduction. This work, inspired by [1–3], considers roving agents' numerical characterization, challenging ad hoc pervasive and trustworthy networks. Agents are autonomous, moving, and intelligent software structures capable to play a sensitive role in advanced monitoring, computation and protection systems. Intrusion detection systems (IDS) [1] are addressed particularly. They appear as complementary mean to the ordinary cryptographic protection tools of computers and networks. Such IDS use software agent based monitoring and data collection, watching the inside processes of a computer, registering LOG files of application software systems, sniffing and recording communication protocols. Watching the whole network behavior they are better suited to warn approaching attacks and malfunctioning. Data mining agents (DMA) and Data fusion agents (DFA) are examples of information integration tools in networks [2]. In large networks, moreover when its structure is not predefined such as wireless sensor networks [3] it is natural to consider independent, randomly roving agents, requiring that they are able to collect enough information in total, mining the necessary knowledge about the intrusion. This framework is studied in [2], which prove formulas for the number of DMA sufficient to monitor the given size areas of networks. The formula received is complex and impractical because of their use of nested sums by different parameters. Our work tends to prove simple estimates for the same numerical characteristics of WSN.

Roving Agents Model. DMA roams around randomly in a network and acquires environmental information. It is lightweight using simplest mining algorithms. DFA is for integration of DMA set actions. DFA may act as an intrusion

* E-mail: hakob.aslanyan@gmail.ru

detection tool and then its power depends on information collected by DMA in network.

Given a network N of n nodes v_1, v_2, \dots, v_n . Some fixed amount of information \mathfrak{I}_i is allocated at node v_i . There are k DMA a_1, a_2, \dots, a_k . Each agent visits exactly m different nodes and obtains the unique information content in each such node. DMA pass all collected information to DFA. Denote by $P_k(n, m, t)$ the probability that DFA contains exactly t information blocks of network nodes when k agents randomly visit m of n nodes each. The formula for $P_k(n, m, t)$ proven in [2] looks as:

$$P_k(n, m, t) = \binom{n}{m}^{-(k-1)} \sum_{m_2, m_3, \dots, m_{k-1}=0}^m \left\{ \binom{m}{m_2} \binom{n-m}{m-m_2} \binom{2m-m_2}{m_3} \binom{n-2m+m_2}{m-m_3} \dots \times \right. \\ \times \dots \left. \binom{(k-2)m-m_2-\dots-m_{k-2}}{m_{k-1}} \binom{n-(k-2)m+m_2+m_{k-2}}{m-m_{k-1}} \right\} \times \\ \times \left. \binom{(k-1)m-m_2-\dots-m_{k-1}}{km-t-m_2-\dots-m_{k-1}} \binom{n-(k-1)m+m_2+m_{k-1}}{t-(k-1)m+m_2+\dots+m_{k-1}} \right\}, k \geq 4. \quad (1)$$

For smaller k formulas, given in [2], look similar. Of course, these formulas are unobservable and simplifications or approximations are of interest. By this same reason [2], proving the formulas, considers computer simulation to understand the typical numbers of agents necessary to retrieve the required information in network. Modifications of “exactly t ” condition in agent distribution scheme are also important to be considered.

Coverage Characterization of Roving Agents. Given the set $N = \{v_1, \dots, v_n\}$ of nodes and S_1, \dots, S_k be k arbitrary subsets of N , of size m , $m \leq n$, visited correspondingly by the k agents. Consider a probability distribution scheme over the N , and suppose that m -subsets S_j are equiprobable and independent. Having in total C_n^m m -subsets the probability of one of them equals $1/C_n^m$. We study the probabilistic characteristics of the union $\bigcup_{i=1}^k S_i$ and of its size. In particular, the proba-

bility $p_t = p\left(\left|\bigcup_{i=1}^k S_i\right|=t\right)$ that the union of those subsets contains exactly t elements.

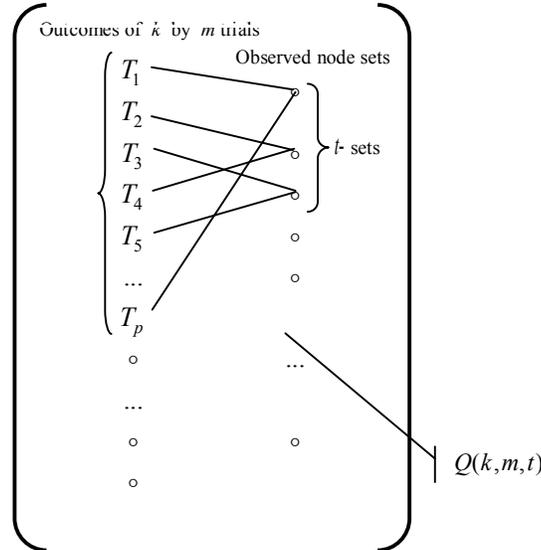
A matrix $A^{k \times n} = \{a_{ij}\}$, where $a_{ij} = \begin{cases} 1, & v_j \in S_i, \\ 0, & \text{otherwise,} \end{cases}$ corresponds to collection

of subsets S_1, \dots, S_k of N nodes. As each S_j contains exactly m elements, each row of $A^{k \times n}$ contains m 1s and $n-m$ 0s. If $\left|\bigcup_{i=1}^k S_i\right|=t$, then there are t columns of

A , which contain at least one 1 and $n-t$ columns, which don't contain 1. The number of $k \times n$ matrixes with m ones on each row and with exactly $n-t$ columns with no 1 is $C_n^t Q(k, m, t)$, where $Q(k, m, t)$ is number of $k \times t$ matrixes with m ones on each row and at least one 1 on each column.

Alternatively, let us consider the following schematic presentation of roving agents' distribution (see [4, 5]). Left column vertices in the Scheme 1 contain all arrangements T_1, T_2, \dots of k agents roving by C_n^m m -node-subsets (ordered collections of k m -node-subsets).

From combinatorial perspective agents and nodes are distinguishable but



Scheme 1. Agent sets distribution in terms of trials and node sets.

m -node-subsets are considered as usual sets – different elements and no ordering. Total number of arrangements is $(C_n^m)^k$. Part of these arrangements covers exactly t nodes. Let these be vertices T_1, T_2, \dots, T_p . We want to compute the unknown number p . Right side column vertices correspond to all subsets of node set N and part of these sets are of size t . In our experiment node subset sizes may take values from m to $\min(km, n)$.

We draw an edge between an arrangement and a node subset, which is covered by that arrangement. Each arrangement

is incident to exactly one edge (and subset). Each t -subset appears in different arrangements and this number say $Q(k, m, t)$, is common for all t -subsets.

$Q(k, m, t)$ can be calculated by inclusion-exclusion principle. We use the matrix model for arrangements. First, over a $k \times t$ -matrix we take the whole set of unconstrained arrangements as all matrices with m 1s on rows, then we remove from this all arrangements, when at least 1 column is initially filled with 0 (such matrices do not obey the conditions we require), then add arrangements with at least 2 empty columns, etc. The formula representation of related quantities is

$$Q(k, m, t) = (C_t^m)^k - C_t^1 (C_{t-1}^m)^k + C_t^2 (C_{t-2}^m)^k - \dots + (-1)^{t-m} (C_t^{t-m})^k = \sum_{i=0}^{t-m} (-1)^i C_t^i (C_{t-i}^m)^k.$$

We have proven

$$\textbf{Theorem 1. } P_k(n, m, t) = \frac{C_n^t \sum_{i=0}^{t-m} (-1)^i C_t^i (C_{t-i}^m)^k}{(C_n^m)^k}.$$

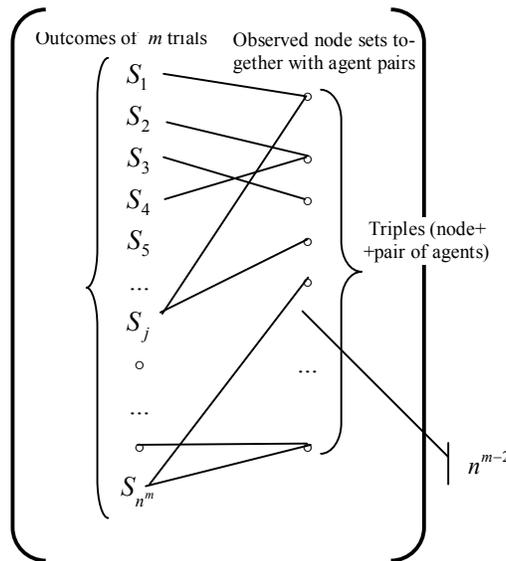
We receive a real simplification of formula (1). The formula received is still complex, but it might be approximated and the applied Markov inequality may give asymptotic estimates of t -subset probabilities. The mean value of subset size t takes the form

$$\sum_{t=m}^{\min(km,n)} t P_k(n, m, t) = \sum_{t=m}^{\min(km,n)} \frac{t C_n^t \sum_{i=0}^{t-m} (-1)^i C_t^i (C_{t-i}^m)^k}{(C_n^m)^k}.$$

On Node Repetition Limitations in An Agent Roving Scheme. Consider the scene of random distribution of m agents over the n WSN nodes (here we do not consider k agents but m agents, and each individual agent visits exactly one node). Agents are dropped over the node set one by one, independently, and with equal probabilities for nodes. Allocating all m agents we receive a collection of nodes visited by agents, probably with multiple agents that visited the same node.

Total number of different allocations is n^m . Among these: 1 node allocations number is n (all agents visit the same node); 2 node allocations, they are $C_n^2(2^m - 2)$; and the larger sets are $n(n-1)\dots(n-m+1)$ m -sets, when agents are distributed in all different nodes. We need in frequencies of allocation sizes, when at least 2 agents are allocated at the same node (sizes from 1 to $m-1$), or complementary, the share of allocations with all different nodes.

One of approaches of determining typical cases in distributions is when Markov (or Chebyshev) inequality is applied. In this way we consider a Scheme 2 similar to Scheme 1 to compute the means of number of allocated nodes in random



Scheme 2. Agents distribution on WSN node sets.

distribution of m agents over the n WSN nodes.

Thus, number of right side vertices equals to $n C_m^2$. Edges are connecting an allocation to a node with the given pair of agents it contains. Similarly to the above case we compute that

$$M(v_{n,m}) = \frac{n C_m^2 n^{m-2}}{n^m} = \frac{C_m^2}{n}.$$

Apply Markov inequality $\Pr\{v_{n,m} \geq \varepsilon\} \leq \frac{M(v_{n,m})}{\varepsilon}$.

Take $\varepsilon = 1$, then C_m^2 / n is the upper estimate of probability of repeating agents at nodes. If $C_m^2 / n \rightarrow 0$ with $n, m \rightarrow \infty$,

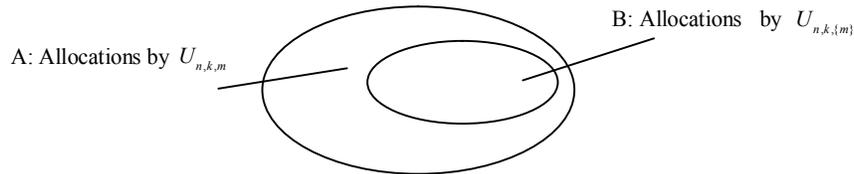
then we receive that almost all allocations consist of all different agents at nodes.

Comparison of Agent Allocation Schemes. In this point we will define and consider two basic probability distributions related to each other (Scheme 3).

- Basic distribution $U_{n,k,\{m\}}$ is formed by k independent consecutive allocations of m -node-subsets over the WSN area of n nodes. $(C_n^m)^k$. Outcomes of trials are ordered collections of m -subsets of WSN nodes. These collections may cover all node subsets of sizes from m to $\min(km, n)$.

• Second distribution $U_{n,k,m}$, which has to be compared with the basic one $U_{n,k,\{m\}}$ introduced above, consists of k consecutive and independent stages; each stage allocates m elements consecutively and independently over the WSN area of n nodes. Outcomes of these trials are all n^{km} ordered collections of nodes. These collections may cover all node subsets of sizes from 1 to $\min(km, n)$.

In one individual stage of $U_{n,k,m}$ we have $m!$ orderings of a single allocation of m -subset of one step of $U_{n,k,\{m\}}$. This is to be taken into account comparing $U_{n,k,\{m\}}$ and $U_{n,k,m}$. This difference can also be seen comparing the one stage outcomes of $U_{n,k,\{m\}}$ and $U_{n,k,m}$. Represent C_n^m of model $U_{n,k,\{m\}}$ as $\frac{n!}{m!(n-m)!} = \frac{n(n-1)\dots(n-m+1)}{m!}$. Numerator of last ratio is the counterpart of n^m of model $U_{n,k,m}$, and $m!$ is the coefficient being mentioned about. Comparing $U_{n,k,\{m\}}$ and $U_{n,k,m}$, first we note that outcomes of $U_{n,k,\{m\}}$ are part of outcomes of $U_{n,k,m}$, and hence they may have larger probabilities.



Scheme 3. Allocations by $U_{n,k,\{m\}}$ and $U_{n,k,m}$.

Consider the probability p_j of an event, that in $U_{n,k,m}$, in stage j all allocated m elements are different. Then, $P = p_1 \cdot p_2 \cdot \dots \cdot p_k$ is the probability that in all k stages allocated m elements are different. In different stages allocations may intersect. Outcomes of $U_{n,k,\{m\}}$ multiplied with this probabilities are equal to probabilities of $U_{n,k,m}$, part B of intersection of outcomes. p_j was estimated in previous point as a value tending to 1 asymptotically. We may extend this proposition to the entire value P . Formally we use the property that probability of events union is less or equal the sum of event probabilities:

$$\Pr\{(v_{n,m} \geq \varepsilon | q=1) \vee (v_{n,m} \geq \varepsilon | q=2) \vee \dots \vee (v_{n,m} \geq \varepsilon | q=k)\} \leq k \Pr\{v_{n,m} \geq \varepsilon\} \leq \frac{kM(v_{n,m})}{\varepsilon}.$$

Then the final condition (upper estimate) sufficient for repetition probability tending to zero is $kC_m^2 / n \rightarrow 0$ with $n, m, k \rightarrow \infty$. The sufficient condition for allocation of all m agents in all k consecutive stages to different nodes $km^2 / n \rightarrow 0$ is naturally acceptable in WSN, which has very large nodes sets as a rule.

Final picture is: part B allocations appear in $U_{n,k,m}$ with probability P tending to 1; relative probability distribution among the elements of B is identical

in $U_{n,k,\{m\}}$ and $U_{n,k,m}$; event probability in model $U_{n,k,\{m\}}$ is not less than in $U_{n,k,m}$ multiplied by P ; probabilities of t -subset allocations under the model $U_{n,k,m}$ have formulas similar to the ones for model $U_{n,k,\{m\}}$ considered above.

If $R(k,m,t)$ denotes the number of t -node allocations in model $U_{n,k,m}$, then the formal representation of $R(k,m,t)$ similar to the formula for $Q(k,m,t)$ considered above can be achieved by the same inclusion/exclusion method:

$$R(k,m,t) = t^{mk} - C_t^1(t-1)^{mk} + C_t^2(t-2)^{mk} - \dots + (-1)^{t-1}(t-1)^{mk} = \sum_{i=0}^{t-1} (-1)^i C_t^i (t-i)^{mk}.$$

On this basis we formulate

Theorem 2. If $\frac{kC_m^2}{n} \rightarrow 0$ with $n, m, k \rightarrow \infty$, then $\frac{C_n^t Q(k,m,t)}{(C_n^m)^k} P \leq \frac{C_n^t R(k,m,t)}{n^{km}}$ with $P \rightarrow 1$.

Finally, we note that $R(k,m,t)$ has equivalent presentation in terms of second kind Stirling numbers $S(N, K) = \frac{1}{K!} \sum_{j=0}^K (-1)^j C_K^j (K-j)^N$. Here we used the fact, that allocation of k consecutive and independent stages of m elements over the WSN area of n nodes is equivalent to allocation of km elements over that area. Note a difference between the formulas for $Q(k,m,t)$ and $R(k,m,t)$ – that is summation limits. In case of $R(k,m,t)$ formally we may add the zero term for $i=t$, and then we receive $R(k,m,t) = t! S(mk, t)$, which is the final postulation of this paper.

Conclusion. WSN and software agent systems are important application technique for many areas. Being hard algorithmically and complex in model level these systems require special economy regimes, and this is concerned in knowing the minimal requirements and maximum effect, when resource is limited. In randomly roving agents model, which is considered above, it is shown that appearing probabilities are equivalently presented in terms of combinatorial Stirling numbers and due to known asymptotic formulas for these numbers [6, 7], this allows to adopt the monitoring regime in an optimal way.

Received 21.01.2010

REFERENCES

1. **Krugel C., Toth T. and Kirda E.** SPARTA, A Mobile Agent Based Intrusion Detection System, First International IFIP TC-11 WG 11.4 Working Conference on Network Security, November 26–27, 2001, Leuven, Belgium.
2. **Moskowitz I.S., Kang M.H., Chang Li Wu and Longdon G.E.** Randomly Roving Agents for Intrusion Detection. Center for High Assurance Computer Systems. Tech. report 5540-TM/02/003, Naval Research Laboratory, Washington D.C., 2001, 16 p.
3. **Brandes U. and Erlebach T.** (Eds.) Network Analysis, LNCS 3418. Berlin, Heidelberg: Springer-Verlag, 2005, 471 p.
4. **Glagolev V.** Problemy Kibernetiki, 1967, № 19, p. 75–94 (in Russian).

5. **Aslanyan H.** Greedy Set Cover Estimations, Int. Conference “Computer Science and Information Technologies”, 22–26 Sept. 2003, Yer., p. 143–144.
6. **Medvedev Yu.I., Ivchenko G.I.** Theory of Probability and Applications, 1965, v. X, № 1, p. 151–156.
7. **Chelluri R., Richmond L.B., Temme N.M.** Asymptotic Estimates for Generalized Stirling Numbers. Report MAS-R9923, ISSN 1386-3703, CWI, 1999, Amsterdam, Netherlands.

Հ. Լ. Ասլանյան

Պատահականորեն թափառող ագենտների քանակական բնութագրում

Աշխատանքում ուսումնասիրվել են անլար սենսորային ցանցերում (WSN) պատահականորեն թափառող ագենտների քանակական բնութագրերը: Բացի հայտնի բանաձևերի պարզեցումներից, ցույց է տրվել, որ հիմնական ագենտային մոդելը հավանականային մակարդակում համարժեք է համանման մի առավել պարզ մոդելի: Այնուհետև ստացվել են հաճախականությունների համարժեք բանաձևեր, որոնք ներկայացված են կոմբինատորիկայում հայտնի Ստիռլինգի երկրորդ սեռի թվերի միջոցով: Ստիռլինգի թվերը խորապես ուսումնասիրված են և նրանց համար հայտնի է տարբեր գնահատականների մի ամբողջ շարք:

А. Л. Асланян.

Количественная характеристика случайно блуждающих агентов

В работе исследованы количественные характеристики случайно блуждающих агентов в беспроводных сенсорных сетях (WSN). Помимо упрощенной формул, известных из публикаций, показана эквивалентность базовой агентной модели более простой модели на вероятностном уровне. Далее выведены эквивалентные формулы для частот в терминах известных комбинаторных чисел Стирлинга второго рода. Числа Стирлинга хорошо изучены и для них известен целый ряд различных оценок.