Review of Research on Network Flow Watermarking Techniques

Hui Chen[1,2], Qianmu Li[1,2(✉)], Shunmei Meng[1,2], Haiyuan Shen[1,2], Kunjin Liu[1,2], and Huaqiu Long[3]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
qianmu@njust.edu.cn
[2] Jiangsu Zhongtian Technology Co., Ltd., Nantong 226463, China
[3] Intelligent Manufacturing Department, Wuyi University, Jiangmen 529020, China

Abstract. In cloud environment, a framework for cross-domain collaborative tracking could find intruders hidden behind autonomous domains by linking these autonomous domains effectively. The autonomous domain in framework could select the appropriate intrusion tracking technology to implement intra-domain tracking according to its own operating rules and communication characteristics. As an active traffic analysis technology, network flow watermarking technology could accurately locate the real positions of intruders hidden behind intermediate hosts (stepping stones) and anonymous communication systems. Furthermore, it has many advantages such as high precise rate, low false alarm rate, short observation time and so on. For these advantages and its high efficiency of intra-domain tracking, it has become the hot spot in academe research in recent years. Therefore, this paper did the following work: (1) research on network flow watermarking technology; (2) conclude the implementation framework of network flow watermarking technology; (3) analyze the principles and implementation processes of several mainstream network flow watermarking schemes; (4) analyze threats to network flow watermarking.

Keywords: Network flow watermarking technology · Intra-domain tracking · Threats to network flow watermarking

1 Mainstream Network Flow Watermarking Techniques

Network flow [1] is a sequence of unidirectional data packets or frames transmitted between any different nodes in the network for a period of time. It is also known as communication flow or packet flow. A unidirectional network flow could be uniquely represented by a tuple made up of the following eight elements: source IP address, destination IP address, source port number, service type, destination port number, protocol number, input interface and output interface. The network flow association determines the communication relationship by detecting the correlation between network flow of sender and receiver, and then implements intrusion tracing.

X. Zhang et al. (Eds.): CloudComp 2019/SmartGift 2019, LNICST 322, pp. 519–528, 2020.
https://doi.org/10.1007/978-3-030-48513-9_41

The closest watermarking embedder to the attacker is responsible for the generation of original watermark signal, watermark coding and watermark modulation. Its concrete process is: First, the watermarking embedder uses random number generator SNG to generate an original watermark signal with a specified length, and encodes it. Then embedder modulates the passing network flow according to a specified strategy of watermarking implementation. After that the original watermarking signal generated is carried by the network flow. Finally, the identifier of network flow and the original watermark signal are stored in a tracking database. In general, there are two deployment schemes of the watermark embedder. One is embedding a specific watermark into the network flow of response message by using NAT or router near the target server, and another is applying a watermark signal to intra-domain network flow in boundaries of autonomous domains.

The watermarking detector is responsible for demodulation, decoding and watermarking similarity comparison of the received network stream. The work flow of detector is mainly as follows: When the network flow carrying watermark signal is transmitted to the watermark detector located near a victim host by network, intermediate hosts (stepping stones) or anonymous communication systems, the watermarking detector first records the characteristics of the received network flow. And then it demodulates network flow to obtain coded signals of watermark according to the watermark parameters shared with the watermark embedder. After getting the final recovered watermark signals by decoding coded signals, the detector could use comparison functions to obtain the similarity between the final recovered watermark signals and the original watermark, such as cosine similarity or Hamming distance. Finally, the similarity is compared with the preestablished thresholds. If the similarity is greater than the given threshold, it is considered that there is an association between sender and recipient of network flow. In other words, the communication relationship between sender and recipient of the network flow could be confirmed.

There is a set of dynamically changing hidden watermark parameters shared between the watermarking embedder and the watermarking detector when the network flow watermarking scheme is implemented. The watermark detector needs to use the same watermark parameters as the embedder to complete the extraction of watermark signal. Watermark signal won't be extracted if parameters are different, which prevents the watermark information from being acquired and attacked by attackers. It means that the robustness and privacy of watermark information are enhanced.

The generation of original watermark signal, selection of watermark carrier and methods of watermark embedding will all affect the effect of tracking when network flow watermarking techniques are used. This section introduced the principles and implementation steps of several existing mainstream network watermarking techniques.

1.1 Network Flow Watermarking Technique Based on Flow Rate

The core idea of network flow watermarking technique based on flow rate is to embed the watermark signal by adjusting the flow rate in different time periods after selecting different time periods in the target network flow duration. The DSSS scheme proposed by Yu et al. [2] combines direct sequence spread spectrum technology with watermark based on flow rate. It not only improves the capacity of the watermark, but also enhances

the ability of tracking multiple attack traffic in parallel. Its implementation framework is shown in Fig. 1:

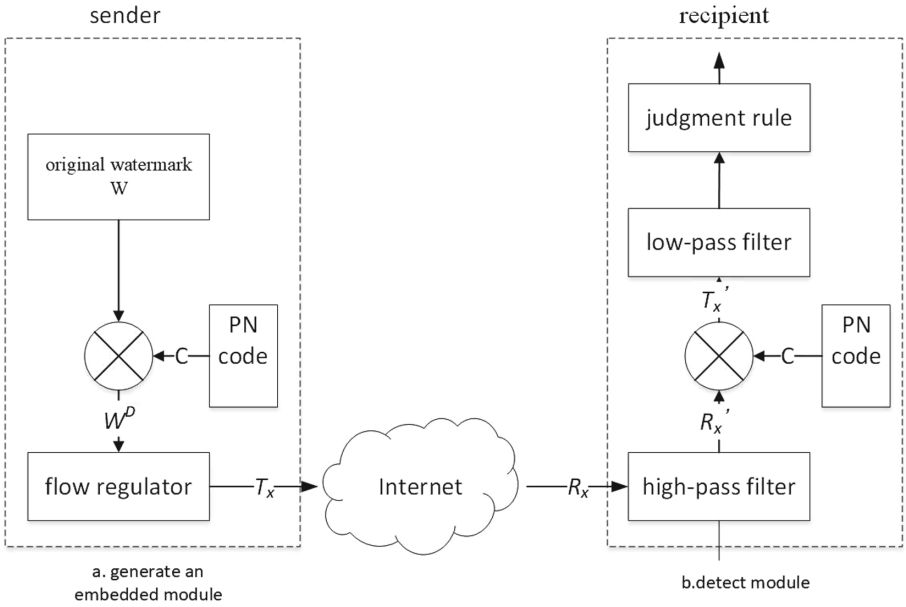

Fig. 1. Implementation framework of DSSS scheme

First, a binary original watermark signal of a specified length $W = \{w_1, w_2, \ldots, w_n\}, |W| = n$ is generated, then a watermark information bit W^D could be gotten after spreading W by DSSS. W^D is expressed by formula 1. And spreading W need set PN Code $C = \{c_1, c_2, \ldots, c_m\}, |C| = m$ in advance. w_i^D is the corresponding watermark signal of w_i after spreading original watermark W. w_i^D is expressed by formula 2.

$$W^D = W \times C = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \times \begin{pmatrix} c_1 \cdots c_m \end{pmatrix} = \begin{pmatrix} w_1^D \\ \vdots \\ w_n^D \end{pmatrix} \tag{1}$$

$$w_i^D = \begin{pmatrix} w_i\, c_1 \cdots w_i c_m \end{pmatrix} \tag{2}$$

After W^D is obtained, each watermark signal of W^D is embedded in the original network flow according to formula 3. In formula 3, S represents the original flow rate of the network flow, T_x represents the flow rate after the watermark signal is modulated, $w_i c_j$ represents the j-th code element of the i-th spread spectrum watermark signal in W^D. And I_{PN} is the duration of modulating each code element.

$$T_x = w_i c_j A + S \tag{3}$$

Let rate of receiving network flow be R_x, which is shown in formula 4. In this formula, z is noise. The watermarking detector first removes DC component S in R_x through high-pass filter, then obtains R'_x which is shown in formula 5.

$$R_x = w_i c_j A + S + z \tag{4}$$

$$R'_x \approx w_i c_j A + z \tag{5}$$

After that, the receiver dispreads R'_x by formula 6 and obtains T'_x. Finally, the original watermark signal could be recovered after noise zC is eliminated through low-pass filter.

$$T'_x = w_i c_j AC + zC \tag{6}$$

1.2 Network Flow Watermarking Technique Based on Inter-packet Delay

For a unidirectional network flow containing n data packets, t_i denotes the time when the i-th packet P_i in the network flow arrives at the current host, and t'_i denotes the time when P_i leaves the current host. For the i-th and j-th packets in the network flow, the inter-packet delay between the arrival time of P_i and P_j is defined as AIPD (arrive inter-packet delay), and the inter-packet delay between the departure time of P_i and P_j is defined as DIPD (departure inter-packet delay). In general, AIPD is used as a carrier for modulating watermarks. So IPD in this paper refers specifically to AIPD.

$$AIPD_{i,j} = t_i - t_j \tag{7}$$

$$DIPD_{i,j} = t'_i - t'_j \tag{8}$$

The watermarking scheme based on inter-packet delay selects the interval delay of several pairs of data packets in the network flow as the watermark carrier, and embeds a watermark signal by adjusting the size of one or more sets of IPD means. Wang et al. [3] proposed an improved IPD scheme when tracking anonymous VOIP telephony traffic. The schematic diagram of its principle is shown in Fig. 2:

Firstly, $A = \{P_1, \ldots, P_{2r}\}$ are denoted by 2r packets selected from the network flow, and $B = \{P_{1+d}, \ldots, P_{2r+d}\}$ are denoted by 2r packets selected in increments of d. For A and B, P_i is a packet, and $|A| = |B| = 2r$. The packets in A and B are mapped one by one to obtain 2r packet pairs $\langle P_1, P_{1+d}\rangle, \ldots, \langle P_{2r}, P_{2r+d}\rangle$, in which r represents redundancy.

The IPDs between each pair of packets in the 2r packet pairs are calculated, and then they are divided into 2 groups, which are recorded as IPD^1 and IPD^2($|IPD^1| = |IPD^2| = r$). $\overline{Y_{r,d}}$ is the mean difference for IPD^1 and IPD^2. It could be calculated by formula 9, 10.

$$Y_{r,d} = \frac{\left(ipd_k^1 - ipd_k^2\right)}{2}, k = 1, 2, \ldots, r \tag{9}$$

$$\overline{Y_{r,d}} = \frac{1}{r}\sum\nolimits_{k=1}^{r} Y_{r,d} \tag{10}$$

The watermark signal is embedded according to the mean difference $\overline{Y_{r,d}}$. And it is denoted as w, $w \in \{-1, 1\}$. The embedding principle is as follows:

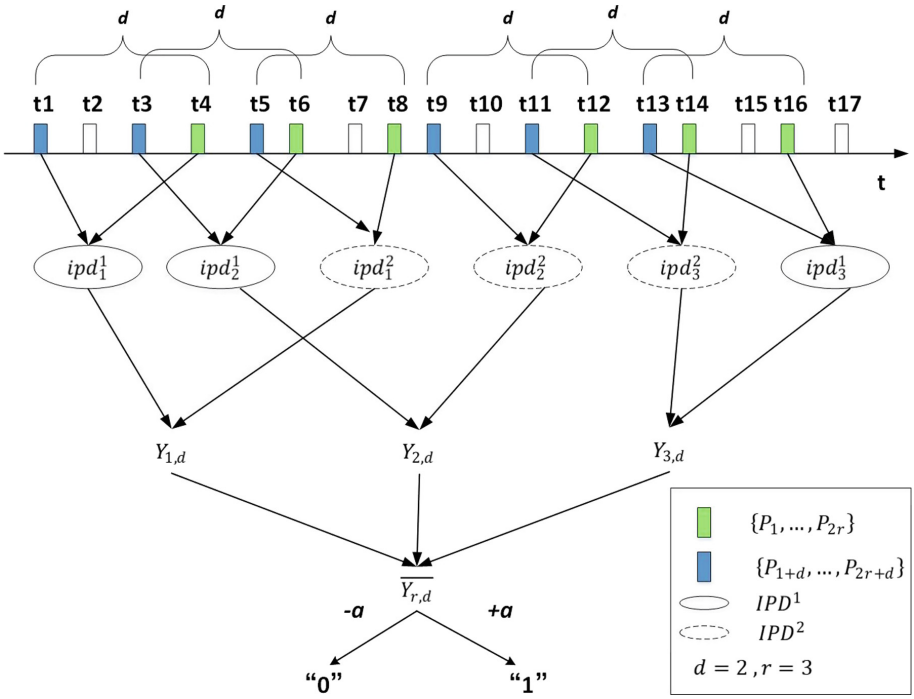

Fig. 2. Example of IPD allocation in network flow watermarking technique based on inter-packet delay

- If $\overline{Y_{r,d}} < 0, w = -1$, IPD^1 and IPD^2 don't need to be adjusted.
- If $\overline{Y_{r,d}} < 0, w = 1$, the IPD of IPD^1 will be increased and the IPD of IPD^2 will be decreased to ensure $\overline{Y_{r,d}} > 0$. The corresponding operation is to delay packets from group B in IPD^1 and packets from group A in IPD^2.
- If $\overline{Y_{r,d}} \geq 0, w = -1$, the IPD of IPD^2 will be increased and the IPD of IPD^1 will be decreased to ensure $\overline{Y_{r,d}} < 0$. The corresponding operation is to delay packets from group A in IPD^1 and packets from group B in IPD^2.
- If $\overline{Y_{r,d}} \geq 0, w = 1$, IPD^1 and IPD^2 don't need to be adjusted.

When the watermark detector receives the network flow, IPD^1 and IPD^2 could be obtained by using the same parameters and random strategies as the watermark embedder. And then $\overline{Y_{r,d}}$ is calculated by formula 9, 10. Finally, the watermark signal W' could be recovered according to formula 11.

$$W' = \begin{cases} -1, \overline{Y_{r,d}} < 0 \\ 1, \overline{Y_{r,d}} \geq 0 \end{cases} \tag{11}$$

1.3 Interval Centroid Based Network Flow Watermarking Technique

Network flow is a sequence of unidirectional data packets or frames transmitted between any different nodes in the network for a period of time. Its duration is divided into n

equal time periods in units of T, which are time slots: $I_0, I_1, \ldots, I_{n-1}$. There are m data packet packets $P_1, P_2, \ldots P_{\mathrm{m}}$ in I_{i}, and $t_{\mathrm{i}}(i = 1, 2, \ldots, m)$ indicates the timestamp of each packet arrival. If t_0 denotes the start time of I_{i}, $\Delta t_i = (t_i - t_0) mod T$ indicates the offset time of P_{i} relative to t_0. The centroid of I_{i} is defined as:

$$\mathrm{C}(I_{\mathrm{i}}) = \frac{1}{m} \sum\nolimits_{i=1}^{m} \Delta t_i \tag{12}$$

Wang et al. [3] proved that if X is used as a divisor to perform a modulo operation on a random variable Y that is much larger than X, the result will approximately obey uniform distribution on [0, X]. It means that if there are enough packets in I_{i}, Δt_i will approximately obeys uniform distribution on [0, X], and the mathematical expectation of Δt_i will be stabilized at a invariant $T/2$ ($E(\Delta t_i) = T/2$). This invariant could be used as a stable watermark carrier.

Because the watermarking scheme based on centroid of time slot uses the centroid of the network slot as carrier, the watermark signal is embedded by changing the centroid value of the time slot. Wang et al. proposed an interval centroid based watermarking (ICBW) scheme when tracking low-rate anonymous communication systems. Its principle is as follows:

As Fig. 3 shows, according to the predefined parameters, the duration of the network flow is divided into 2n time slots ($I_0, I_1, \ldots, I_{2n-1}$) in units of T. These time slots are equally divided into group A and B by random selection function. Group A and B are denoted as $I_k^A (k = 0, 1, \ldots, n-1)$, $I_k^B (k = 0, 1, \ldots, n-1)$, $|A| = |B| = n$. And then an original watermark signal W whose length is l could be generated ($W = \{w_1, w_2, \ldots, w_l\}$, $|W| = l$).

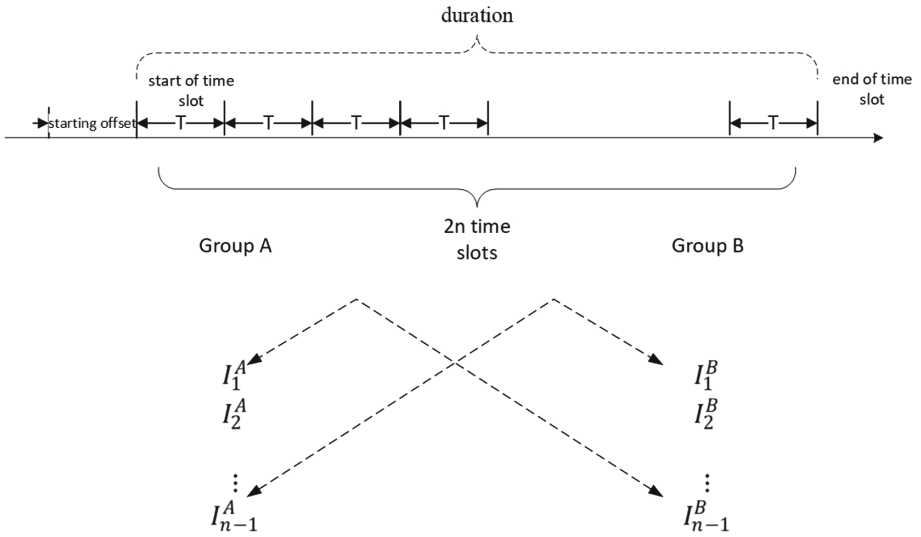

Fig. 3. Example of time slot selection and allocation in ICBW scheme

After that r time slots are randomly extracted from group A and group B to embed the watermark signal w_i. Each time slot is denoted as $I_{i,j}^{A}(j = 0, 1, \ldots, r-1)$ or $I_{i,j}^{B}(j = 0, 1, \ldots, r-1)$, and r $(r = n/l)$ represents redundancy. $I_{i,j}^{A}$ represents the j-th time slot with w_i in group A, and $I_{i,j}^{B}$ represents the j-th time slot with w_i in group B. A_i and B_i (the combined slot centroids of $I_{i,j}^{A}$ and $I_{i,j}^{B}$) are calculated by formula 13, 14. $N_{i,j}^{A}$ indicates the number of packets in $I_{i,j}^{A}$ and $N_{i,j}^{B}$ indicates the number of packets in $I_{i,j}^{B}$. $Cent\left(I_{i,j}^{A}\right)$ is the centroid of $I_{i,j}^{A}$ and $Cent\left(I_{i,j}^{B}\right)$ is the centroid of $I_{i,j}^{B}$. N_i^A and N_i^B indicate the total number of packets with w_i in group A and B respectively.

$$A_i = \frac{\sum_{j=0}^{r-1}\left[N_{i,j}^{A} Cent\left(I_{i,j}^{A}\right)\right]}{\sum_{j=0}^{r-1} N_{i,j}^{A}} = \frac{\sum_{j=0}^{r-1}\left[N_{i,j}^{A} Cent\left(I_{i,j}^{A}\right)\right]}{N_i^A} \tag{13}$$

$$B_i = \frac{\sum_{j=0}^{r-1}\left[N_{i,j}^{B} Cent\left(I_{i,j}^{B}\right)\right]}{\sum_{j=0}^{r-1} N_{i,j}^{B}} = \frac{\sum_{j=0}^{r-1}\left[N_{i,j}^{B} Cent\left(I_{i,j}^{B}\right)\right]}{N_i^B} \tag{14}$$

$$Y_i = A_i - B_i \tag{15}$$

After Y_i is calculated by formula 15, w_i is embedded by adjusting Y_i. The specific embedding method is as follows:

- If $w_i = 1$, A_i and B_i will be adjusted to ensure $Y_i \approx \alpha/2(\alpha$ is modulation amplitude). The adjustment scheme is to compress the time of packet in $I_{i,j}^{A}$ according to formula 16, which could increase its slot centroid. In formula 16, $\Delta t_{i,j,k}$ indicates the offset of the k-th packet of $I_{i,j}$ relative to the start time slot, and $\Delta t'_{i,j,k}$ is the calculated offset value. And the corresponding time slots in group B remain unchanged.

$$\Delta t'_{i,j,k} = \alpha + \frac{T - \alpha}{T} \Delta t_{i,j,k} \tag{16}$$

- If $w_i = -1$, A_i and B_i will be adjusted to ensure $Y_i \approx -\alpha/2(\alpha$ is modulation amplitude). The adjustment scheme is to compress the time of packet in $I_{i,j}^{B}$ according to formula 16, which could increase its slot centroid. And the corresponding time slots in group A remain unchanged.

When the watermark detector receives the network flow carrying watermark signals, $2n$ time slots are obtained according to the same parameters and divided into group A and B by using the same random allocation strategy S. After that, A_i, B_i and Y_i could be calculated by formula 13–15.

2 Threats to Network Flow Watermarking Technologies

Network flow watermarking technology embeds watermark signals in network flow by actively modifying the traffic characteristics of the network flow. So these characteristics

also become targets for attackers to detect and decipher watermarks. This section mainly introduces the following attack methods:

(1) Watermark Attack Based on Digital Filtering
In reference [4], an intrusion tracking technology which use flow rate to modulate watermark in anonymous networks was studied. Based on the research, a watermark detection method based on digital filtering was proposed. The method first intercepts the network flow with a watermark online, and add an identifier to it. The method then uses the Bayesian classifier for offline training to analyze frequency domain of its characteristic. Finally, the method can identify whether the network flow carries a watermark. Experiments show that the method can effectively detect network flow watermarking technique based on flow rate and effectively enhance the Wireless Mix Network's ability to resist this technique.

(2) Attack Based on Time Analysis
By analyzing the watermarking scheme based on inter-packet delay, Peng et al. [5] proposed an attack method based on the analysis of IPD between adjacent intermediate hosts(stepping stones). The attack method uses expectation maximization algorithm to estimate the proportion of quantization step to packet delay, and uses the Bayesian classification rule to identify the packet carrying an identifier. Peng et al. gave specific measures for watermark recovery or removal in four cases. These measures allow an attacker to remove a watermark from a chain of stepping stones or to copy a watermark to a chain without stepping stones in some cases. After evaluating the detection rate of watermarking, the false positive rate, and the minimum number of delayed packets, the attack method they proposing can detect the watermark signal by the sequential probability ratio. It can be used to real-timely detect on whether the network flow carries a watermark.

(3) Multi-flow Attack
Kiyavash et al. [6] studied several timing-based flow watermarking schemes (e.g. IBW, ICBW, DSSS). They found that modulating watermark will produce a long empty time segment without any packets arriving because these schemes divide the duration of the network flow into different time slots to embed the watermark signal. The long empty time segment can facilitate attackers to discover the existence of watermark. Attackers often align multiple watermarks carrying network flows and conduct a multi-flow attack (MFA). MFA can be used to detect the watermark signal carried in the network flow, and even recover the watermark parameter to remove the watermark in the network flow. In addition, it can also be used to detect network flows carrying different watermark signals.

(4) Attack Based on MSAC (Mean-Square Autocorrelation)
In the spread spectrum watermarking scheme like DSSS-W, the direct sequence spread spectrum of the watermark signal is often performed by using the pseudo noise code (PN code). The method can significantly reduce the interference of network jittering to the efficiency of watermarking tracking, and enhance the stability of watermark signals. It has been widely adopted. However, the network flow using the same PN code has

strong autocorrelation. Jia et al. [7] defined the value of MSAC and determined whether there is a watermark in the network flow by calculating value of MSAC. The scheme that they proposed is called attack based on MSAC. Compared with MFA, attack based on MSAC only needs to compare multiple segments of a data stream to detect the existence of watermark signals, which is simpler and more efficient.

(5) DSSS Watermark Removal Attack Based on TCP Flow Control Mechanism
In the network flow watermarking based on inter-packet delay, embedding watermark signals will cause periodic alternating of high throughput and low throughput. And DSSS makes the periodic alternating more visible. Luo et al. [8] proposed LZPL attack. The attack first detects the abnormal sequence by locating the period of low throughput in the network flow, and removes the watermark signal with spread spectrum modulated in the network stream through the TCP flow control mechanism. It does not require the support of routers and relay nodes. So it is easy for implementation.

3 Conclusion

In cloud environment, a framework for cross-domain collaborative tracking could find intruders hidden behind autonomous domains by linking these autonomous domains effectively. The autonomous domain in framework could select the appropriate intrusion tracking technology to implement intra-domain tracking according to its own operating rules and communication characteristics. As an active traffic analysis technology, network flow watermarking technology could accurately locate the real positions of intruders hidden behind intermediate hosts (stepping stones) and anonymous communication systems. Furthermore, it has many advantages such as high precise rate, low false alarm rate, short observation time and so on.

Acknowledgments. This work was supported in part by the Fundamental Research Funds for the Central Universities (30918012204), Military Common Information System Equipment Preresearch Special Technology Project (315075701), 2019 Industrial Internet Innovation and Development Project from Ministry of Industry and Information Technology of China, 2018 Jiangsu Province Major Technical Research Project "Information Security Simulation System", Shanghai Aerospace Science and Technology Innovation Fund (SAST2018-103).

References

1. Callado, A.: A survey on internet traffic identification. IEEE Commun. Surv. Tutorials **11**(3), 37–52 (2009)
2. Yu, W., Fu, X.: DSSS-based flow marking technique for invisible traceback. In: 2007 IEEE Symposium on Security and Privacy, Berkeley, CA, USA, pp. 18–32. IEEE (2007)
3. Wang, X., Chen, S.: Tracking anonymous peer-to-peer VoIP calls on the internet. In: 12th ACM Conference on Computer and Communications Security, pp. 81–91. ACM, Newyork (2005)
4. Li, Q.: Safety risk monitoring of cyber-physical power systems based on ensemble learning algorithm. IEEE Access **7**, 24788–24805 (2019)

5. Peng, P., Ning, P.: On the secrecy of timing-based active watermarking trace-back techniques. In: 2006 IEEE Symposium on Security and Privacy, Berkeley/Oakland, CA, USA, pp. 315–349. IEEE (2006)
6. Kiyavash, N., Houmansadr, A.: Multi-flow attacks against network flow watermarking schemes. In: 17th USENIX Security Symposium, pp. 307–320. USENIX Association, Berkeley (2008)
7. Jia, W., Tso, F.P.: Blind detection of spread spectrum flow watermarks. In: 28th IEEE International Conference on Computer Communications, Rio de Janeiro, Brazil, pp. 2195–2203. IEEE (2009)
8. Luo, X., Zhang, J.: On the secrecy of spread-spectrum flow watermarks. In: Gritzalis, D., Preneel, B. (eds.) European Conference on Research in Computer Security 2010. LNCS, vol. 6345, pp. 232–248. Springer, Heidelberg (2010)