[0007] With reference to the first possible implementation manner or the second possible implementation manner of the first aspect, in a third possible implementation manner, the rule further includes that a distance between the at least two points p_e and p_f is 1 U.

[0008] With reference to the first aspect, or any one of the first to third possible implementation manners of the first aspect, in a fourth possible implementation manner, the determining whether at least a part of data in the window $W_{iz}[p_{iz}-A_z, p_{iz}+B_z]$ meets a preset condition C_z specifically includes:

determining, by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z, p_{iz}+B_z]$ meets the preset condition C_z.

[0009] With reference to the fourth possible implementation manner of the first aspect, in a fifth possible implementation manner, the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z, p_{iz}+B_z]$ meets the preset condition C_z is specifically: determining, by using a hash function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z, p_{iz}+B_z]$ meets the preset condition C_z.

[0010] With reference to the first aspect, or any one of the first to fifth possible implementation manners of the first aspect, in a sixth possible implementation manner, when the at least a part of data in the window $W_{iz}[p_{iz}-A_z, p_{iz}+B_z]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the point p_{iz} along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[p_{ic}-A_c, p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[p_{iz}-A_z, p_{iz}+B_z]$ or a left boundary of a window $W_{ic}[p_{ic}-A_c, p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point falls within a range of the window $W_{iz}[p_{iz}-A_z, p_{iz}+B_z]$, where the point p_{ic} determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

[0011] With reference to the fourth possible implementation manner of the first aspect, in a seventh possible implementation manner, the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z, p_{iz}+B_z]$ meets the preset condition C_z specifically includes:

specifically: determining, by using a hash function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z.

[0018] With reference to the second aspect, or any one of the first to fifth possible implementation manners of the second aspect, in a sixth possible implementation manner, when the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the current potential dividing point k_i along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[k_i - A_z, k_i + B_z]$ or a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point falls within a range of the window $W_{iz}[k_i - A_z, k_i + B_z]$, where the window $W_{ic}[k_i - A_c, k_i + B_c]$ determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the rule.

[0019] With reference to the fourth possible implementation manner of the second aspect, in a seventh possible implementation manner, the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z specifically includes:

selecting F bytes in the window $W_{iz}[k_i - A_z, k_i + B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, where F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, where when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, where $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as:

p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window W_{iz} [p_{iz} - A_z , p_{iz} + B_z] or a left boundary of a window W_{ic} [p_{ic} - A_c , p_{ic} + B_c] corresponding to a point p_{ic} that is determined for the new potential dividing point falls within a range of the window W_{iz} [p_{iz} - A_z , p_{iz} + B_z], where the point p_{ic} determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

[0027] With reference to the fourth possible implementation manner of the third aspect, in a seventh possible implementation manner, the determining, by the central processing unit by using a random function, whether the at least a part of data in the window W_{iz} [p_{iz} - A_z , p_{iz} + B_z] meets the preset condition C_z specifically includes:

selecting F bytes in the window W_{iz} [p_{iz} - A_z , p_{iz} + B_z], and using the F bytes repeatedly H times to obtain F*H bytes in total, where F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, where when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, where $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + \ldots + V_{am,8} * h_{m,8}$, S_{a1} , S_{a2} , ..., and S_{aF*H} are obtained

[0030] With reference to the first possible implementation manner or the second possible implementation manner of the fourth aspect, in a third possible implementation manner, the rule further includes A_e-1= A_f and B_e +1= B_f.

[0031] With reference to the fourth aspect, or any one of the first to third possible implementation manners, in a fourth possible implementation manner, the central processing unit is specifically configured to:

determine, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z.

[0032] With reference to the fourth possible implementation manner of the fourth aspect, in a fifth possible implementation manner, the central processing unit is specifically configured to determine, by using a hash function, whether the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z.

[0033] With reference to the fourth aspect, or any one of the first to fifth possible implementation manners, in a sixth possible implementation manner, when the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the current potential dividing point k_i along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[k_i - A_c,\ k_i + B_c]$ that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ or a left boundary of a window $W_{ic}[k_i - A_c,\ k_i + B_c]$ that is determined for the new potential dividing point falls within a range of the window $W_{iz}[k_i - A_z,\ k_i + B_z]$, where the window $W_{ic}[k_i - A_c,\ k_i + B_c]$ determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the rule.

[0034] With reference to the fourth possible implementation manner of the fourth aspect, in a seventh possible implementation manner, the determining, by the central processing unit by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z specifically includes:

selecting F bytes in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, where F≥1, each byte is formed by 8 bits, which are denoted as

$a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, where when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, where $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + \ldots + V_{am,8} * h_{m,8}$, S_{a1}, S_{a2}, ..., and S_{aF*H} are obtained in a same way, a quantity K of values greater than 0 among S_{a1}, S_{a2}, ..., and S_{aF*H} is counted, and when K is an even number, the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z.

[0035] According to a fifth aspect, an embodiment of the present invention provides a server for searching for a data stream dividing point, where a rule is preset on the server, where the rule is: for a potential dividing point k, determining M points p_x, a window $W_x[p_x - A_x, p_x + B_x]$ corresponding to the point p_x, and a preset condition C_x corresponding to the window $W_x[p_x - A_x, p_x + B_x]$, where x indicates consecutive natural numbers from 1 to M, M≥2, and A_x and B_x are integers; and

the server includes: a determining ~~processing~~ unit, configured to perform step (a):

(a) determining a point p_{iz} and a window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ corresponding to the point p_{iz} for a current potential dividing point k_i according to the rule, where i and z are integers, and 1≤z≤M; and

to: when the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ does not meet the preset condition C_z, skip the N minimum units U for searching for a data stream dividing point from the point p_{iz} along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, where the determining unit performs step (a) for the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ or a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ ~~that is determined for the new potential dividing point~~ falls within a range of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$, where the point p_{ic} ~~window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$~~ determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

[0042] With reference to the fourth possible implementation manner of the fifth aspect, in a seventh possible implementation manner, that the judging processing unit is specifically configured to determine, by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z specifically includes:

selecting F bytes in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, where F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, where when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, where $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as:

[0045] With reference to the first possible implementation manner of the sixth aspect, in a second possible implementation manner, the rule further includes that A_e and A_f are positive integers.

[0046] With reference to the first possible implementation manner or the second possible implementation manner of the sixth aspect, in a third possible implementation manner, the rule further includes A_e-1= A_f and B_e +1= B_f.

[0047] With reference to the sixth aspect, or any one of the first to third possible implementation manners, in a fourth possible implementation manner, the judging and processing unit is specifically configured to:

determine, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z.

[0048] With reference to the fourth possible implementation manner of the sixth aspect, in a fifth possible implementation manner, the judging and processing unit is specifically configured to determine, by using a hash function, whether the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z.

[0049] With reference to the sixth aspect, or any one of the first to fifth possible implementation manners, in a sixth possible implementation manner, the judging and processing unit is configured to: when the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ does not meet the preset condition C_z, skip the N minimum units U for searching for a data stream dividing point from the current potential dividing point k_i along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, the determining unit performs step (a) for the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[k_i - A_c,\ k_i + B_c]$ that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ or a left boundary of a window $W_{ic}[k_i - A_c,\ k_i + B_c]$ that is determined for the new potential dividing point falls within a range of the window $W_{iz}[k_i - A_z,\ k_i + B_z]$, where the window $W_{ic}[k_i - A_c,\ k_i + B_c]$ determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the rule.

[0050] With reference to the fourth possible implementation manner of the sixth aspect, in a seventh possible implementation manner, that the judging and processing unit determines, by using

in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z.

[0056] With reference to the fourth possible implementation manner of the seventh aspect, in a fifth possible implementation manner, the determining, by the server by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z specifically includes:

determining, by the server by using a hash function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z.

[0057] With reference to the seventh aspect, or any one of the first to fifth possible implementation manners of the seventh aspect, in a sixth possible implementation manner, when the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the point p_{iz} along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ or a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point falls within a range of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$, where the point p_{ic} determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to ~~a the data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

[0058] With reference to the fourth possible implementation manner of the seventh aspect, in a seventh possible implementation manner, the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z specifically includes:

selecting F bytes in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, where F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, where

potential dividing point k_i along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and accordin to the rule, a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[k_i - A_z, k_i + B_z]$ or a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point falls within a range of the window $W_{iz}[k_i - A_z, k_i + B_z]$, where the window $W_{ic}[k_i - A_c, k_i + B_c]$ determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the rule.

[0066] With reference to the fourth possible implementation manner of the eighth aspect, in a seventh possible implementation manner, the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z specifically includes:

selecting F bytes in the window $W_{iz}[k_i - A_z, k_i + B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, where F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, where when $a_{m,n}=1$, $V_{am,n}=1$, and when $a_{m,n}=0$, $V_{am,n}=-1$, where $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of

[0101] The determining whether at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets a preset condition C_z specifically includes:

determining, by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z.

[0102] The determining, by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z is specifically: determining, by using a hash function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z.

[0103] When the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the point p_{iz} along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ or a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point falls within a range of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$, where the point p_{ic} determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

[0104] In this embodiment of the present invention, a data stream dividing point is searched for by determining whether at least a part of data in a window of M windows meets a preset condition, and when the at least a part of data in the window does not meet the preset condition, a length of N*U is skipped, where N*U is not greater than $\| B_z \| + \max_x(\| A_x \| + \| (k_i - p_{ix}) \|)$, so as to obtain a next potential dividing point, thereby improving efficiency of searching for a data stream dividing point.

[0105] In a process of eliminating duplicate data, to ensure an even size of a data chunk, a size of an average data chunk (also referred to as an average chunk) is considered. That is, while limits on a size of a minimum data chunk and a size of a maximum data chunk are met, the size of the average data chunk is determined to ensure an even size of an obtained data chunk. A probability (represented as P(n)) of finding a data stream dividing point depends on two factors, that is, the

potential dividing point k_i and the potential dividing point k_j all fall within a determining range, in this implementation manner, it may be ensured that a left boundary of the window $W_{j11}[p_{j11}$ -169, $p_{j11}]$ coincides with a right boundary p_{i5} of $W_{i5}[p_{i5}$-169, $p_{i5}]$ or falls within a range of $W_{i5}[p_{i5}$-169, $p_{i5}]$, where the point p_{j11} determined for the potential dividing point k_j is a point ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the potential dividing point k_j according to the rule. Therefore, within such a limit, when the at least a part of data in $W_{i5}[p_{i5}$-169, $p_{i5}]$ does not meet the preset condition C_5, a distance for skipping from p_{i5} along the direction of searching for a data stream dividing point is not greater than $\| B_5 \| + \max_x(\| A_x \| + \| (k_i - p_{ix}) \|)$, where M=11, and 11*U is not greater than $\max_x(\| A_x \| + \| (k_i - p_{ix}) \|)$, and therefore, a distance for skipping from p_{i5} along the direction of searching for a data stream dividing point is not greater than 179. It is determined whether at least a part of data in $W_{j1}[p_{j1}$-169, $p_{j1}]$ meets the preset condition C_1, it is determined whether at least a part of data in $W_{j2}[p_{j2}$-169, $p_{j2}]$ meets the preset condition C_2, it is determined whether at least a part of data in $W_{j3}[p_{j3}$-169, $p_{j3}]$ meets the preset condition C_3, it is determined whether at least a part of data in $W_{j4}[p_{j4}$-169, $p_{j4}]$ meets the preset condition C_4, it is determined whether at least a part of data in $W_{j5}[p_{j5}$-169, $p_{j5}]$ meets the preset condition C_5, it is determined whether at least a part of data in $W_{j6}[p_{j6}$-169, $p_{j6}]$ meets the preset condition C_6, it is determined whether at least a part of data in $W_{j7}[p_{j7}$-169, $p_{j7}]$ meets the preset condition C_7, it is determined whether at least a part of data in $W_{j8}[p_{j8}$-169, $p_{j8}]$ meets the preset condition C_8, it is determined whether at least a part of data in $W_{j9}[p_{j9}$-169, $p_{j9}]$ meets the preset condition C_9, it is determined whether at least a part of data in $W_{j10}[p_{j10}$-169, $p_{j10}]$ meets the preset condition C_{10}, and it is determined whether at least a part of data in $W_{j11}[p_{j11}$-169, $p_{j11}]$ meets the preset condition C_{11}. Certainly, in this embodiment of the present invention, the rule is also followed when it is determined whether the potential dividing point k_a is a data stream dividing point, specific implementation is not described again, and reference may be made to the description of determining the potential dividing point k_i. When it is determined that the at least a part of data in the window

corresponding to the point p_{i5}, and $C_5=C_{11}$; therefore, for the potential dividing point k_i, when it is determined that the at least a part of data in $W_{i5}[\ p_{i5}\text{-}169,\ p_{i5}]$ does not meet the preset condition C_5, the potential dividing point k_g obtained by skipping 10 bytes from the point p_{i5} along the direction of searching for a data stream dividing point still does not meet a condition of serving as a data stream dividing point. Therefore, if repeated calculation exists when 10 bytes are skipped from the point p_{i5} along the direction of searching for a data stream dividing point, then repeated calculation can be reduced and efficiency is higher when 11 bytes are skipped from the point p_{i5} along the direction of searching for a data stream dividing point. Therefore, a speed of searching for a data stream dividing point is increased. When ~~in a preset specification,~~ a probability that at least a part of data in the window $W_x[\ p_x\text{-}A_x,\ p_x+B_x]$ corresponding to the point p_x that is determined according to the preset rule meets the preset condition C_x is 1/2, that is, a probability of executing skipping is 1/2, at most 179 bytes can be skipped each time.

[0108] In this implementation manner, a preset rule is: for a potential dividing point k, determining 11 points p_x, a window $W_x[\ p_x\text{-}A_x,\ p_x+B_x]$ corresponding to the point p_x, and a preset condition C_x corresponding to the window $W_x[\ p_x\text{-}A_x,\ p_x+B_x]$, where x indicates consecutive natural numbers from 1 to 11 separately, where a probability that at least a part of data in the window $W_x[\ p_x\text{-}A_x,\ p_x+B_x]$ corresponding to the point p_x meets the preset condition is 1/2, and P(n) can be calculated by using the two factors, that is, the quantity of points p_x and the probability that at least a part of data in the window $W_x[\ p_x\text{-}A_x,\ p_x+B_x]$ corresponding to the point p_x meets the preset condition. Moreover, $A_1=A_2=A_3=A_4=A_5=A_6=A_7=A_8=A_9=A_{10}=A_{11}=169$, $B_1=B_2=B_3=B_4=B_5=B_6=B_7=B_8=B_9=B_{10}=B_{11}=0$, and $C_1=C_2=C_3=C_4=C_5=C_6=C_7=C_8=C_9=C_{10}=C_{11}$, where a distance between p_x and the potential dividing point k is d_x bytes. Specifically, a distance between p_1 and the potential dividing point k is 0 byte, a distance between p_2 and k is 1 byte, a distance between p_3 and k is 2 bytes, a distance between p_4 and k is 3 bytes, a distance between p_5 and k is 4 bytes, a distance between p_6 and k is 5 bytes, a distance between p_7 and k is 6 bytes, a distance between p_8 and k is 7 bytes, a distance between p_9 and k is 8 bytes, a distance between p_{10} and k is 9 bytes, a distance between p_{11} and k is 10 bytes, and relative to the potential dividing point k, all p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}, and p_{11} are in a direction opposite to the direction of searching for a data stream dividing point. Therefore, whether the potential dividing point k is a data stream

therefore the matrix R is used: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{255,1} & h_{255,2} & \cdots & h_{255,8} \end{pmatrix}$, and random numbers of an mth row of the matrix V_b and the mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{bm} = V_{bm,1} * h_{m,1} + V_{bm,2} * h_{m,2} + ... + V_{bm,8} * h_{m,8}$. S_{b1}, S_{b2}, ..., and S_{b255} are obtained according to the method, and a quantity K of values meeting a specific condition (being greater than 0 is used as an example here) among S_{b1}, S_{b2}, ..., and S_{b255} is counted. Because the matrix R follows normal distribution, S_{bm} still follows normal distribution as the matrix R does. According to the probability theory, the probability that a random number in normal distribution is greater than 0 is 1/2; a probability that each value among S_{b1}, S_{b2}, ..., and S_{b255} is greater than 0 is 1/2, and therefore, K meets binomial distribution: $P(k=n) = C_{255}^{n}(\frac{1}{2})^{n}(\frac{1}{2})^{255-n} = C_{255}^{n}(\frac{1}{2})^{255}$. According to a counting result, it is determined whether the quantity K of values greater than 0 among S_{b1}, S_{b2}, ..., and S_{b255} is an even number; the probability that a random number in binomial distribution is an even number is 1/2, and therefore, K meets a condition with a probability of 1/2. When K is an even number, it indicates that the at least a part of data in W_{i2} [p_{i2}-169, p_{i2}] meets the preset condition C_2. When K is an odd number, it indicates that the at least a part of data in W_{i2} [p_{i2}-169, p_{i2}] does not meet the preset condition C_2. C_2 here refers to that the quantity K, which is obtained according to the foregoing manner, of values greater than 0 among S_{b1}, S_{b2}, ..., and S_{b255} is an even number. In the implementation manner shown in FIG. 3, the at least a part of data in W_{i2} [p_{i2}-169, p_{i2}] meets the preset condition C_2.

[0121] Therefore, as shown in FIG. 16, "▧" represents 1 byte selected when it is determined whether at least a part of data in the window W_{i3} [p_{i3}-169, p_{i3}] meets a preset condition C_3, and there are 42 bytes between two adjacent selected bytes. Selected 5 bytes of data are used repeatedly 51 times to obtain 255 bytes in total, so as to increase randomness. Then, the method for determining whether at least a part of data in the windows W_{i1} [p_{i1}-169, p_{i1}] and W_{i2} [p_{i2}-169, p_{i2}] meets a preset condition is used to determine whether the at least a part of data in W_{i3} [p_{i3} -169, p_{i3}] meets the preset condition C_3.In the implementation manner shown in FIG. 5, the at least a part of data in W_{i3} [p_{i3}-169, p_{i3}] meets the preset condition. As shown in FIG. 16, "▨" represents 1

byte selected when it is determined whether at least a part of data in the window $W_{i4}[p_{i4}\text{-}169, p_{i4}]$ meets a preset condition C_4, and there are 42 bytes between two adjacent selected bytes. Selected 5 bytes of data are used repeatedly 51 times to obtain 255 bytes in total, so as to increase randomness. Then, the method for determining whether at least a part of data in windows $W_{i1}[p_{i1}\text{-}169, p_{i1}]$, $W_{i2}[p_{i2}\text{-}169, p_{i2}]$, and $W_{i3}[p_{i3}\text{-}169, p_{i3}]$ meets a preset condition is used to determine whether the at least a part of data in $W_{i4}[p_{i4}\text{-}169, p_{i4}]$ meets the preset condition C_4. In the implementation manner shown in FIG. 5, the at least a part of data in $W_{i4}[p_{i4}\text{-}169, p_{i4}]$ meets the preset condition C_4. As shown in FIG. 16, "⊞" represents 1 byte selected when it is determined whether at least a part of data in the window $W_{i5}[p_{i5}\text{-}169, p_{i5}]$ meets a preset condition C_5, and there are 42 bytes between two adjacent selected bytes. Selected 5 bytes of data are used repeatedly 51 times to obtain 255 bytes in total, so as to increase randomness. Then, the method for determining whether at least a part of data in the windows $W_{i1}[p_{i1}\text{-}169, p_{i1}]$, $W_{i2}[p_{i2}\text{-}169, p_{i2}]$, $W_{i3}[p_{i3}\text{-}169, p_{i3}]$, and $W_{i4}[p_{i4}\text{-}169, p_{i4}]$ meets a preset condition is used to determine whether the at least a part of data in $W_{i5}[p_{i5}\text{-}169, p_{i5}]$ meets the preset condition C_5. In the implementation manner shown in FIG. 5, the at least a part of data in $W_{i5}[p_{i5}\text{-}169, p_{i5}]$ does not meet the preset condition C_5.

[0122] When the at least a part of data in $W_{i5}[p_{i5}\text{-}169, p_{i5}]$ does not meet the preset condition C_5, 11 bytes are skipped from a point p_{i5} along a direction of searching for a data stream dividing point, and a next potential dividing point k_j is obtained at an end position of an 11th byte. As shown in FIG. 6, according to the rule preset on the deduplication server 103, a point p_{j1} and a window $W_{j1}[p_{j1}\text{-}169, p_{j1}]$ corresponding to the point p_{j1} are determined for the potential dividing point k_j. A manner of determining whether at least a part of data in the window $W_{j1}[p_{j1}\text{-}169, p_{j1}]$ meets the preset condition C_1 is the same as a manner of determining whether at least a part of data in the window $W_{i1}[p_{i1}\text{-}169, p_{i1}]$ meets the preset condition C_1. Therefore, as shown in FIG. 17, W_{j1} represents the window $W_{j1}[p_{j1}\text{-}169, p_{j1}]$, and 5 bytes are selected to determine whether the at least a part of data in $W_{j1}[p_{j1}\text{-}169, p_{j1}]$ meets the preset condition C_1. In FIG. 17, "■" represents 1 selected byte, and there are 42 bytes between two adjacent selected bytes. The selected 5 bytes of data are used repeatedly 51 times to obtain 255 bytes in total, so as to increase randomness. Each byte thereof is formed by 8 bits, which are denoted as $a_{m,1}'$, ..., and

least a part of data in $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ meets the preset condition C_1. When K is an odd number, it indicates that the at least a part of data in $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ does not meet the preset condition C_1.

[0123] A manner of determining whether at least a part of data in $W_{i2}[p_{i2}\text{-}169,\ p_{i2}]$ meets the preset condition C_2 is the same as a manner of determining whether at least a part of data in $W_{j2}[p_{j2}\text{-}169,\ p_{j2}]$ meets the preset condition C_2. Therefore, as shown in FIG. 17, "▩" represents 1 byte selected when it is determined whether the at least a part of data in the window $W_{j2}[p_{j2}\text{-}169,\ p_{j2}]$ meets the preset condition C_2, and there are 42 bytes between two adjacent selected bytes. Selected 5 bytes of data are used repeatedly 51 times to obtain 255 bytes in total, so as to increase randomness. Each byte thereof is formed by 8 bits, which are denoted as $b_{m,1}'$, ..., and $b_{m,8}'$, representing the 1st bit to the 8th bit of an mth byte in the 255 bytes, and therefore, bits corresponding to the 255 bytes may be represented as: $\begin{pmatrix} b_{1,1}' & b_{1,2}' & \cdots & b_{1,8}' \\ b_{2,1}' & b_{2,2}' & \cdots & b_{2,8}' \\ \vdots & \vdots & \vdots & \vdots \\ b_{255,1}' & b_{255,2}' & \cdots & b_{255,8}' \end{pmatrix}$, where when $b_{m,n}'=1$, $V_{bm,n}'=1$, and when $b_{m,n}'=0$, $V_{bm,n}'=-1$, where $b_{m,n}'$ represents any one of $b_{m,1}'$, ..., and $b_{m,8}'$, and a matrix V_b' is obtained according to a conversion relationship between $b_{m,n}'$ and $V_{bm,n}'$ from the bits corresponding to the 255 bytes, and may be represented as: $\begin{pmatrix} V_{b1,1} & V_{b1,2} & \cdots & V_{b1,8} \\ V_{b2,1} & V_{b2,2} & \cdots & V_{b2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{b255,1} & V_{b255,2} & \cdots & V_{b255,8} \end{pmatrix}$. Manners of determining whether at least a part of data in windows $W_2[p_2\text{-}169,\ p_2]$ and $W_2[q_2\text{-}169,\ q_2]$ meets a preset condition are the same, and therefore the matrix R is still used: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{255,1} & h_{255,2} & \cdots & h_{255,8} \end{pmatrix}$. Random numbers of an mth row of the matrix V_b' and the mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{bm}'=V_{bm,1}'*h_{m,1}+V_{bm,2}'*h_{m,2}+\ldots+V_{bm,8}'*h_{m,8}$. S_{b1}', S_{b2}', ..., and S_{b255}' are obtained according to the method, and a quantity K of values meeting a specific condition (being greater than 0 is used as an example here) among S_{b1}', S_{b2}', ..., and S_{b255}' is counted. Because

determined whether at least a part of data in $W_{i1}[\ p_{i1}\text{-}169,\ p_{i1}\ ]$ meets a preset condition C_1. As shown in FIG. 16, W_{i1} represents the window $W_{i1}[\ p_{i1}\text{-}169,\ p_{i1}\ ]$, and 5 bytes are selected to determine whether the at least a part of data in $W_{i1}[\ p_{i1}\text{-}169,\ p_{i1}\ ]$ meets the preset condition C_1. In FIG. 16, "■" represents 1 selected byte, and there are 42 bytes between two adjacent selected bytes "■". In one implementation manner, a hash function is used to calculate the 5 selected bytes, and values obtained by means of calculation by using the hash function are in fixed and even distribution. If a value obtained by means of calculation by using the hash function is an even number, it is determined that the at least a part of data in $W_{i1}[\ p_{i1}\text{-}169,\ p_{i1}\ ]$ meets the preset condition C_1, that is, C_1 represents that a value obtained by means of calculation by using the hash function and according to the foregoing manner is an even number. Therefore, a probability that ~~whether~~ at least a part of data in $W_{i1}[\ p_{i1}\text{-}169, p_{i1}\ ]$ meets a preset condition is 1/2. In the implementation manner shown in FIG. 5, it is determined, by using a hash function, whether at least a part of data in $W_{i2}[\ p_{i2}\text{-}169,\ p_{i2}\ ]$ meets a preset condition C_2, whether at least a part of data in $W_{i3}[\ p_{i3}\text{-}169,\ p_{i3}\ ]$ meets a preset condition C_3, whether at least a part of data in $W_{i4}[\ p_{i4}\text{-}169,\ p_{i4}\ ]$ meets a preset condition C_4, and whether at least a part of data in $W_{i5}[\ p_{i5}\text{-}169,\ p_{i5}\ ]$ meets a preset condition C_5. For specific implementation, reference may be made to the description of the manner of determining, by using a hash function, whether at least a part of data in $W_{i1}[\ p_{i1}\text{-}169,\ p_{i1}\ ]$ meets the preset condition C_1 in the implementation manner shown in FIG. 5, which is not described herein again.

[0125] When the at least a part of data in $W_{i5}[\ p_{i5}\text{-}169,\ p_{i5}\ ]$ does not meet the preset condition C_5, 11 bytes are skipped from a point p_{i5} along a direction of searching for a data stream dividing point, and a current potential dividing point k_j is obtained at an end position of an 11th byte. As shown in FIG. 6, according to the rule preset on the deduplication server 103, a point p_{j1} and a window $W_{j1}[\ p_{j1}\text{-}169,\ p_{j1}\ ]$ corresponding to the point p_{j1} are determined for the potential dividing point k_j. A manner of determining whether at least a part of data in the window $W_{j1}[\ p_{j1}\text{-}169,\ p_{j1}\ ]$ meets the preset condition C_1 is the same as a manner of determining whether at least a part of data in the window $W_{i1}[\ p_{i1}\text{-}169,\ p_{i1}\ ]$ meets the preset condition C_1. Therefore, as shown in FIG. 17, W_{j1} represents the window $W_{j1}[\ p_{j1}\text{-}169,\ p_{j1}\ ]$, and 5 bytes are selected to determine whether the at least a part of data in $W_{j1}[\ p_{j1}\text{-}169,\ p_{j1}\ ]$ meets the preset condition C_1.

In FIG. 17, " " represents 1 selected byte, and there are 42 bytes between two adjacent selected bytes "". The 5 bytes selected from the window $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ are calculated by using a hash function. If an obtained value is an even number, the at least a part of data in $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ meets the preset condition C_1. In FIG. 17, a manner of determining whether at least a part of data in $W_{i2}[p_{i2}\text{-}169,\ p_{i2}]$ meets the preset condition C_2 is the same as a manner of determining whether at least a part of data in $W_{j2}[p_{j2}\text{-}169,\ p_{j2}]$ meets the preset condition C_2. Therefore, as shown in FIG. 17, " " represents 1 byte selected when it is determined whether the at least a part of data in the window $W_{j2}[p_{j2}\text{-}169,\ p_{j2}]$ meets the preset condition C_2, and there are 42 bytes between two adjacent selected bytes "". Selected 5 bytes are calculated by using a hash function. If an obtained value is an even number, the at least a part of data in $W_{j2}[p_{j2}\text{-}169,\ p_{j2}]$ meets the preset condition C_2. In FIG. 17, a manner of determining whether at least a part of data in $W_{i3}[p_{i3}\text{-}169,\ p_{i3}]$ meets the preset condition C_3 is the same as a manner of determining whether at least a part of data in $W_{j3}[p_{j3}\text{-}169,\ p_{j3}]$ meets the preset condition C_3. Therefore, as shown in FIG. 17, " " represents 1 byte selected when it is determined whether the at least a part of data in the window $W_{j3}[p_{j3}\text{-}169,\ p_{j3}]$ meets the preset condition C_3, and there are 42 bytes between two adjacent selected bytes "". Selected 5 bytes are calculated by using a hash function. If an obtained value is an even number, the at least a part of data in $W_{j3}[p_{j3}\text{-}169,\ p_{j3}]$ meets the preset condition C_3. In FIG. 17, a manner of determining whether at least a part of data in $W_{j4}[p_{j4}\text{-}169,\ p_{j4}]$ meets the preset condition C_4 ~~and~~ is the same as a manner of determining whether at least a part of data in the window $W_{i4}[p_{i4}\text{-}169,\ p_{i4}]$ meets the preset condition C_4. Therefore, as shown in FIG. 17, " " represents 1 byte selected when it is determined whether the at least a part of data in the window $W_{j4}[p_{j4}\text{-}169,\ p_{j4}]$ meets the preset condition C_4, and there are 42 bytes between two adjacent selected bytes "". Selected 5 bytes are calculated by using a hash function. If an obtained value is an even number, the at least a part of data in $W_{j4}[p_{j4}\text{-}169,\ p_{j4}]$ meets the preset condition C_4. According to the foregoing method, it is determined whether at least a part of data in $W_{j5}[p_{j5}\text{-}169,\ p_{j5}]$ meets the preset condition C_5, it is determined whether at least a part of data in $W_{j6}[p_{j6}\text{-}169,\ p_{j6}]$ meets a preset condition C_6, it is determined whether at least a part of data

window $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ corresponding to the point p_{j1} are determined for the potential dividing point k_j. A manner of determining whether at least a part of data in the window $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ meets the preset condition C_1 is the same as the manner of determining whether at least a part of data in the window $W_{i1}[p_{i1}\text{-}169,\ p_{i1}]$ meets the preset condition C_1. Therefore, as shown in FIG. 17, W_{j1} represents the window $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$, and ~~to determine whether at least a part of data in $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ meets the preset condition C_1,~~ 5 bytes are selected to determine whether the at least a part of data in $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ meets the preset condition C_1. In FIG. 17, each of bytes "■" whose sequence numbers are 169, 127, 85, 43, and 1 represents 1 selected byte, and there are 42 bytes between two adjacent selected bytes. The bytes "■" whose sequence numbers are 169, 127, 85, 43, and 1 are regarded as 40 sequential bits, which are represented as a_1', a_2', a_3', a_4', ..., and a_{40}' separately. For any a_t' of a_1', a_2', a_3', a_4', ..., and a_{40}', when a_t'=0, V_{at}'=-1, and when a_t'=1, V_{at}'=1. V_{a1}', V_{a2}', V_{a3}', V_{a4}', ..., and V_{a40}' are generated according to a correspondence between a_t' and V_{at}'. The manner of determining whether at least a part of data in the window $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ meets the preset condition C_1 is the same as the manner of determining whether at least a part of data in the window $W_{i1}[p_{i1}\text{-}169,\ p_{i1}]$ meets the preset condition C_1, and therefore, the same random numbers are used: h_1, h_2, h_3, h_4, ..., and h_{40}. $S_a' = V_{a1}' * h_1 + V_{a2}' * h_2 + V_{a3}' * h_3 + V_{a4}' * h_4 + ... + V_{a40}' * h_{40}$. Because h_1, h_2, h_3, h_4, ..., and h_{40} follow normal distribution, S_a' also follows normal distribution. When S_a' is a positive number, the at least a part of data in $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ meets the preset condition C_1; when S_a' is a negative number or 0, the at least a part of data in $W_{j1}[p_{j1}\text{-}169,\ p_{j1}]$ does not meet the preset condition C_1. A probability that S_a' is a positive number is 1/2.

[0137] The manner of determining whether at least a part of data in $W_{i2}[p_{i2}\text{-}169,\ p_{i2}]$ meets the preset condition C_2 is the same as a manner of determining whether at least a part of data in $W_{j2}[p_{j2}\text{-}169,\ p_{j2}]$ meets the preset condition C_2. Therefore, as shown in FIG. 17, "▩" represents 1 byte selected when it is determined whether the at least a part of data in the window $W_{j2}[p_{j2}\text{-}169,\ p_{j2}]$ meets the preset condition C_2, and there are 42 bytes between two adjacent selected bytes. In FIG. 17, selected bytes are represented as sequence numbers 170, 128, 86, 44, and

condition C_z is provided. In this embodiment, it is determined, by using a random function, whether the at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ meets the preset condition C_z. According to a rule preset on a deduplication server 103, a point p_{i1} and a window $W_{i1}[\ p_{i1}-169,\ p_{i1}\ ]$ corresponding to p_{i1} are determined for a potential dividing point k_i, and it is determined whether at least a part of data in $W_{i1}[\ p_{i1}-169,\ p_{i1}\ ]$ meets a preset condition C_1. As shown in FIG. 16, W_{i1} represents the window $W_{i1}[\ p_{i1}-169,\ p_{i1}\ ]$, and 5 bytes are selected to determine whether the at least a part of data in $W_{i1}[\ p_{i1}-169,\ p_{i1}\ ]$ meets the preset condition C_1. In FIG. 16, each of bytes "■" whose sequence numbers are 169, 127, 85, 43, and 1 represents 1 selected byte, and there are 42 bytes between two adjacent selected bytes. Each of the bytes "■" whose sequence numbers are 169, 127, 85, 43, and 1 is converted into 1 decimal number, the range of which is 0-(2^40-1). An even distribution random number generator is used to generate 1 designated value for each decimal number in 0-(2^40-1), and a correspondence R between each decimal number in 0-(2^40-1) and the designated value is recorded. Once being designated, a designated value corresponding to a decimal number remains unchanged, and the designated value follows even distribution. If the designated value is an even number, the at least a part of data in $W_{i1}[\ p_{i1}-169,\ p_{i1}\ ]$ meets the preset condition C_1; if the designated value is an odd number, the at least a part of data in $W_{i1}[\ p_{i1}-169,\ p_{i1}\ ]$ does not meet the preset condition C_1. C_1 represents that the designated value obtained according to the foregoing method is an even number. Because a probability that a random number in even distribution is an even number is 1/2, a probability that the at least a part of data in $W_{i1}[\ p_{i1}-169,\ p_{i1}\ ]$ meets the preset condition C_1 is 1/2. In the implementation manner shown in FIG. 5, by using the same rule, separately, it is determined whether at least a part of data in $W_{i2}[\ p_{i2}-169,\ p_{i2}\ ]$ meets a preset condition C_2, it is determined whether at least a part of data in $W_{i3}[\ p_{i3}-169,\ p_{i3}\ ]$ meets a preset condition C_3, it is determined whether at least a part of data in $W_{i4}[\ p_{i4}-169,\ p_{i4}\ ]$ meets a preset condition C_4, and it is determined whether at least a part of data in $W_{i5}[\ p_{i5}-169,\ p_{i5}\ ]$ meets a preset condition C_5, which are not described herein again.

[0140] When the at least a part of data in $W_{i5}[\ p_{i5}-169,\ p_{i5}\ ]$ does not meet the preset condition C_5, 11 bytes are skipped from a point p_{i5} along a direction of searching for a data stream dividing point, and a current potential dividing point k_j is obtained at an end position of an 11th byte. As shown in FIG. 6, according to the rule preset on the deduplication server 103, a point

C_{10}, and it is determined whether at least a part of data in $W_{j11}[\,p_{j11}\text{-}169,\ p_{j11}\,]$ meets a preset condition C_{11}, which are not described herein again.

[0141] The deduplication server 103 in the embodiment of the present invention shown in FIG. 1 refers to an apparatus that can implement a technical solution described in an embodiment of the present invention, and as shown in FIG. 18, usually includes a central processing unit, a primary memory, and an input/output interface. The central processing unit, the primary memory, and the input/output interface communicate with each other. The main memory stores an executable instruction, and the central processing unit executes the executable instruction stored in the primary memory, so as to perform a specific function, for example, searching for a data stream dividing point described in FIG. 4 to FIG. 17 in the embodiments of the present invention. Therefore, as shown in FIG. 19, according to the embodiments of the present invention shown in FIG. 4 to FIG. 17, for the deduplication server 103, a rule is preset on the deduplication server 103, where the rule is: for a potential dividing point k, determining M points p_x, a window $W_x[\,p_x\text{-}A_x,\ p_x+B_x]$ corresponding to the point p_x, and a preset condition C_x corresponding to the window $W_x[\,p_x\text{-}A_x,\ p_x+B_x]$, where x indicates consecutive natural numbers from 1 to M, M≥2, and A_x and B_x are integers. The deduplication server 103 includes a determining unit 1901 and a judging and processing unit 1902. The determining unit 1901 is configured to perform step (a): (a) determining a point p_{iz} and a window $W_{iz}[\,p_{iz}\text{-}A_z,\ p_{iz}+B_z]$ corresponding to the point p_{iz} for a current potential dividing point k_i according to the rule, where i and z are integers, and 1≤z≤M. The judging and processing unit 1902 is configured to: determine whether at least a part of data in the window $W_{iz}[\,p_{iz}\text{-}A_z,\ p_{iz}+B_z]$ meets a preset condition C_z;

when the at least a part of data in the window $W_{iz}[\,p_{iz}\text{-}A_z,\ p_{iz}+B_z]$ does not meet the preset condition C_z, skip N minimum units U for searching for a data stream dividing point from the point p_{iz} along a direction of searching for a data stream dividing point, where N*U is not greater than $\|\,B_z\,\| + \max_x(\,\|\,A_x\,\| + \|\,(k_i\text{-}p_{ix})\,\|\,)$, so as to obtain a new potential dividing point, where the determining unit performs step (a) for the new potential dividing point; and when at least a part of data in each window $W_{ix}[\,p_{ix}\text{-}A_x,\ p_{ix}+B_x]$ of M windows of the current potential dividing point k_i meets the preset condition C_x, select the current potential dividing point k_i as a data stream dividing point.

[0142] Further, the rule further includes that at least two points p_e and p_f meet conditions

the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + ... + V_{am,8} * h_{m,8}$, S_{a1}, S_{a2}, ..., and S_{aF*H} are obtained in a same way, a quantity K of values greater than 0 among S_{a1}, S_{a2}, ..., and S_{aF*H} is counted, and when K is an even number, the at least a part of data in the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ meets the preset condition C_z.

[0145] Further, the judging and processing unit 1902 is configured to: when the at least a part of data in the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ does not meet the preset condition C_z, skip the N minimum units U for searching for a data stream dividing point from the point p_{iz} along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and the determining unit 1901 performs step (a) for the new potential dividing point, where and according to the rule, a left boundary of a window $W_{ic}[p_{ic} - A_c, p_{ic} + B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ or a left boundary of a window $W_{ic}[p_{ic} - A_c, p_{ic} + B_c]$ ~~that is determined for the new potential dividing point~~ falls within a range of the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$, where the point p_{ic} ~~window $W_{ic}[p_{ic} - A_c, p_{ic} + B_c]$~~ determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to ~~a~~ the ~~data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

[0146] According to the method for searching for a data stream dividing point based on a server in the embodiments of the present invention shown in FIG. 4 to FIG. 17, a point p_{ix} and a window $W_{ix}[p_{ix} - A_x, p_{ix} + B_x]$ corresponding to~~of~~ the point p_{ix} are determined for a potential dividing point k_i, where x indicates consecutive natural numbers from 1 to M separately, and M≥2. It may be determined in parallel whether at least a part of data in each window of M windows meets a preset condition C_x, or it may be determined sequentially whether at least a part of data in windows meets a preset condition, or it may be first determined that at least a part of data in a window $W_{i1}[p_{i1} - A_1, p_{i1} + B_1]$ meets a preset condition C_1, then it is determined that at least a part of data in $W_{i2}[p_{i2} - A_2, p_{i2} + B_2]$ meets a preset condition C_2, until it is determined that at least a part of data in $W_{im}[p_{im} - A_m, p_{im} + B_m]$ meets a preset condition C_m. Other windows in this embodiment are determined in a same manner as the above, which is not described again.

[0147] In addition, according to the embodiments of the present invention shown in FIG. 4 to FIG. 17, in actual applications, a rule is preset on the deduplication server 103, where the rule is: for

the window W_{if}.

[0158] Further, the determining whether at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z specifically includes: determining, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z. Still further, the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z is specifically: determining, by using a hash function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z.

[0159] When the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the current potential dividing point k_i along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point.According to the rule, a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[k_i - A_z, k_i + B_z]$ or a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point falls within a range of the window $W_{iz}[k_i - A_z, k_i + B_z]$, where the window $W_{ic}[k_i - A_c, k_i + B_c]$ determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to ~~a~~ the ~~data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the rule.

[0160] In this embodiment of the present invention, a data stream dividing point is searched for by determining whether at least a part of data in a window of M windows meets a preset condition, and when the at least a part of data in the window does not meet the preset condition, a length of N*U is skipped, where N*U is not greater than $\| B_z \| + \max_x(\| A_x \|)$, so as to obtain a next potential dividing point, thereby improving efficiency of searching for a data stream dividing point.

[0161] In a process of eliminating duplicate data, to ensure an even size of a data chunk, a size of an average data chunk (also referred to as an average chunk) is considered. That is, while limits on a size of a minimum data chunk and a size of a maximum data chunk are met, the size of the average data chunks is determined to ensure an even size of an obtained data chunk. A probability (represented as P(n)) of finding a data stream dividing point depends on two factors, that is, the quantity M of the windows $W_x[k - A_x, k + B_x]$ and a probability that at least a part of data in the

condition C_x corresponding to the window $W_x[k-A_x, k+B_x]$, where x indicates consecutive natural numbers from 1 to 11, and A_x and B_x are integers. A_1=169, B_1=0; A_2=170, B_2=-1; A_3=171, B_3=-2; A_4=172, B_4=-3; A_5=173, B_5=-4; A_6=174, B_6=-5; A_7=175, B_7=-6; A_8=176, B_8=-7; A_9=177, B_9=-8; A_{10}=178, B_{10}=-9; A_{11}=179, B_{11}=-10. $C_1=C_2=C_3=C_4=C_5=C_6=C_7=C_8=C_9=C_{10}=C_{11}$. Therefore, the 11 windows are W_1[k-169, k], W_2[k-170, k-1], W_3[k-171, k-2], W_4[k-172, k-3], W_5[k-173, k-4], W_6[k-174, k-5], W_7[k-175, k-6], W_8[k-176, k-7], W_9[k-177, k-8], W_{10}[k-178, k-9], and W_{11}[k-179, k-10]. k_a is a data stream dividing point, and a direction of searching for a data stream dividing point shown in FIG. 21 is from left to right. After a minimum data chunk of 4 KB is skipped from the data stream dividing point k_a, an end position of the minimum data chunk of 4 KB serves as a next potential dividing point k_i. According to the rule preset ~~for~~ on the deduplication server 103, a window $W_{ix}[k_i-A_x, k_i+B_x]$ is determined for the potential dividing point k_i, and in this embodiment, x indicates consecutive natural numbers from 1 to 11 separately. In the implementation manner shown in FIG. 21, 11 windows are determined for the potential dividing point k_i, and are $W_{i1}[k_i$-169, $k_i]$, $W_{i2}[k_i$-170, k_i-1], $W_{i3}[k_i$-171, k_i-2], $W_{i4}[k_i$-172, k_i-3], $W_{i5}[k_i$-173, k_i-4], $W_{i6}[k_i$-174, k_i-5], $W_{i7}[k_i$-175, k_i-6], $W_{i8}[k_i$-176, k_i-7], $W_{i9}[k_i$-177, k_i-8], $W_{i10}[k_i$-178, k_i-9], and $W_{i11}[k_i$-179, k_i-10] separately. It is determined whether at least a part of data in $W_{i1}[k_i$-169, $k_i]$ meets the preset condition C_1, it is determined whether at least a part of data in $W_{i2}[k_i$-170, k_i-1] meets the preset condition C_2, it is determined whether at least a part of data in $W_{i3}[k_i$-171, k_i-2] meets the preset condition C_3, it is determined whether at least a part of data in $W_{i4}[k_i$-172, k_i-3] meets the preset condition C_4, it is determined whether at least a part of data in $W_{i5}[k_i$-173, k_i-4] meets the preset condition C_5, it is determined whether at least a part of data in $W_{i6}[k_i$-174, k_i-5] meets the preset condition C_6, it is determined whether at least a part of data in $W_{i7}[k_i$-175, k_i-6] meets the preset condition C_7, it is determined whether at least a part of data in $W_{i8}[k_i$-176, k_i-7] meets the preset condition C_8, it is determined whether at least a part of data in $W_{i9}[k_i$-177, k_i-8] meets the preset condition C_9, it is determined whether at least a part of data in $W_{i10}[k_i$-178, k_i-9] meets the preset condition C_{10}, and it is determined whether at least a part of data in $W_{i11}[k_i$-179, k_i-10] meets the preset condition C_{11}. When it is determined that the at least a part of data in

window $W_{j11}[k_j\text{-}179,\ k_j\text{-}10]$ is a window ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of the 11 ~~M~~ windows that are determined for the potential dividing point k_j according to the rule. Therefore, within such a limit, when at least a part of data in the window $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ does not meet the preset condition C_5, a distance for skipping from the potential dividing point k_i along the direction of searching for a data stream dividing point is not greater than $\| B_5 \| + \max_x (\| A_x \|)$. It is determined whether at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1, it is determined whether at least a part of data in $W_{j2}[k_j\text{-}170,\ k_j\text{-}1]$ meets the preset condition C_2, it is determined whether at least a part of data in $W_{j3}[k_j\text{-}171,\ k_j\text{-}2]$ meets the preset condition C_3, it is determined whether at least a part of data in $W_{j4}[k_j\text{-}172,\ k_j\text{-}3]$ meets the preset condition C_4, it is determined whether at least a part of data in $W_{j5}[k_j\text{-}173,\ k_j\text{-}4]$ meets the preset condition C_5, it is determined whether at least a part of data in $W_{j6}[k_j\text{-}174,\ k_j\text{-}5]$ meets the preset condition C_6, it is determined whether at least a part of data in $W_{j7}[k_j\text{-}175,\ k_j\text{-}6]$ meets the preset condition C_7, it is determined whether at least a part of data in $W_{j8}[k_j\text{-}176,\ k_j\text{-}7]$ meets the preset condition C_8, it is determined whether at least a part of data in $W_{j9}[k_j\text{-}177,\ k_j\text{-}8]$ meets the preset condition C_9, it is determined whether at least a part of data in $W_{j10}[k_j\text{-}178,\ k_j\text{-}9]$ meets the preset condition C_{10}, and it is determined whether at least a part of data in $W_{j11}[k_j\text{-}179,\ k_j\text{-}10]$ meets the preset condition C_{11}. When it is determined that the at least a part of data in the window W_{j1} meets the preset condition C_1, the at least a part of data in the window W_{j2} meets the preset condition C_2, the at least a part of data in the window W_{j3} meets the preset condition C_3, the at least a part of data in the window W_{j4} meets the preset condition C_4, the at least a part of data in the window W_{j5} meets the preset condition C_5, the at least a part of data in the window W_{j6} meets the preset condition C_6, the at least a part of data in the window W_{j7} meets the preset condition C_7, the at least a part of data in the window W_{j8} meets the preset condition C_8, the at least a part of data in the window W_{j9} meets the preset condition C_9, the at least a part of data in the window W_{j10} meets the preset condition C_{10}, and the at least a part of data in the window W_{j11} meets the preset condition C_{11}, the current potential dividing point k_j ~~k_i~~ is a data stream dividing point, and data between k_j and

k_a forms 1 data chunk. Meanwhile, a minimum chunk size of 4 KB is skipped in a manner the same as that of k_a to obtain a next potential dividing point, and according to the rule preset on the deduplication server 103, it is determined whether the next potential dividing point is a data stream dividing point. When it is determined that the potential dividing point k_j is not a data stream dividing point, a next potential dividing point is obtained in a manner the same as that of k_i, and it is determined, according to the rule preset on the deduplication server 103 and the foregoing method, whether the next potential dividing point is a data stream dividing point. When no data stream dividing point is found after a set maximum data chunk is exceeded, an end position of the maximum data chunk serves as a forced dividing point.

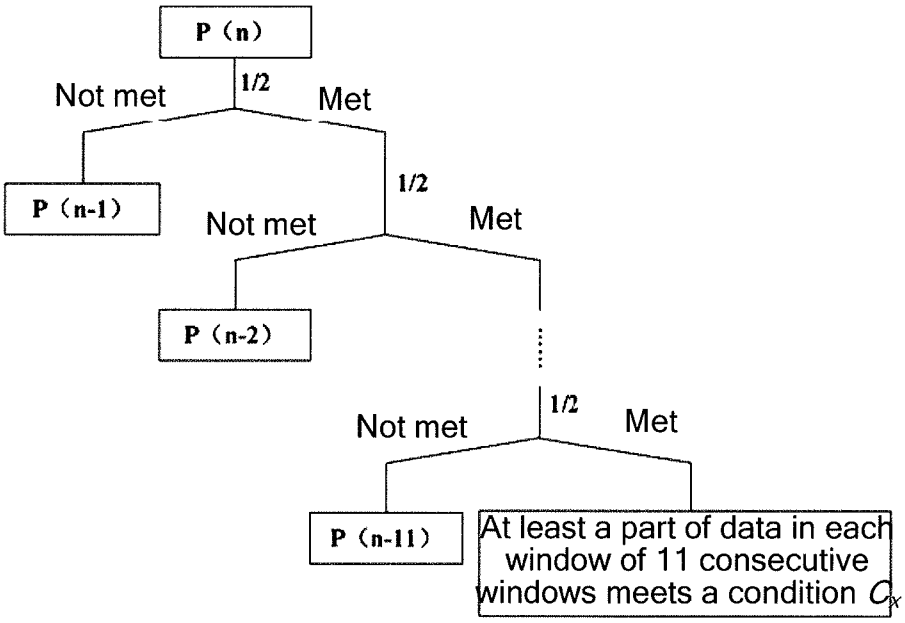
[0163] In the implementation manner shown in FIG. 21, according to the rule preset on the deduplication server 103, starting from determining whether the at least a part of data in $W_{i1}[k_i$ -169, $k_i]$ meets the preset condition C_1, when it is determined the at least a part of data in ~~the at least a part of data in~~ $W_{i1}[k_i$-169, $k_i]$, the at least a part of data in $W_{i2}[k_i$-170, k_i-1], the at least a part of data in $W_{i3}[k_i$-171, k_i-2], and the at least a part of data in $W_{i4}[k_i$-172, k_i-3] meet the preset conditions C_1, C_2, C_3, and C_4 respectively, and it is determined that the at least a part of data in $W_{i5}[k_i$-173, k_i-4] does not meet the preset condition C_5, 6 bytes are skipped from the potential dividing point k_i along the direction of searching for a data stream dividing point, and a new potential dividing point is obtained at an end position of a 6th byte. For differentiation from other potential dividing points, the new potential dividing point is represented as k_g here. According to the rule preset on the deduplication server 103, 11 windows are determined for the potential dividing point k_g, and are $W_{g1}[k_g$-169, $k_g]$, $W_{g2}[k_g$-170, k_g-1], $W_{g3}[k_g$-171, k_g-2], $W_{g4}[k_g$-172, k_g-3], $W_{g5}[k_g$-173, k_g-4], $W_{g6}[k_g$-174, k_g-5], $W_{g7}[k_g$-175, k_g-6], $W_{g8}[k_g$-176, k_g-7], $W_{g9}[k_g$-177, k_g-8], $W_{g10}[k_g$-178, k_g-9], and $W_{g11}[k_g$-179, k_g-10] separately. It is determined whether at least a part of data in $W_{g1}[k_g$-169, $k_g]$ meets the preset condition C_1, it is determined whether at least a part of data in $W_{g2}[k_g$-170, k_g-1] meets the preset condition C_2, it is determined whether at least a part of data in $W_{g3}[k_g$-171, k_g-2] meets the preset condition C_3, it is determined whether at least a part of data in $W_{g4}[k_g$-172, k_g-3] meets the preset condition C_4, it is determined whether at least a part of data in $W_{g5}[k_g$-173, k_g-4] meets the preset condition C_5, it is determined whether at least a part of data in $W_{g6}[k_g$-174,

k_g -5] meets a preset condition C_6, it is determined whether at least a part of data in $W_{g7}[k_g$ -175, k_g -6] meets a preset condition C_7, it is determined whether at least a part of data in $W_{g8}[k_g$ -176, k_g -7] meets a preset condition C_8, it is determined whether at least a part of data in $W_{g9}[k_g$ -177, k_g -8] meets a preset condition C_9, it is determined whether at least a part of data in $W_{g10}[k_g$ -178, k_g -9] meets a preset condition C_{10}, and it is determined whether at least a part of data in $W_{g11}[k_g$ -179, k_g -10] meets a preset condition C_{11}. The window $W_{g11}[k_g$ -179, k_g -10] coincides with the window $W_{i5}[k_i$ -173, k_i -4], and $C_5 = C_{11}$; therefore, when it is determined that the at least a part of data in $W_{i5}[k_i$ -173, k_i -4] does not meet the preset condition C_5, a potential dividing point k_g obtained by skipping T bytes from the potential dividing point k_i along the direction of searching for a data stream dividing point still does not meet a condition for serving as a data stream dividing point. Therefore, if repeated calculation exists when 6 bytes are skipped from the potential dividing point k_i along the direction of searching for a data stream dividing point, then repeated calculation can be reduced and efficiency is higher when 7 bytes are skipped from the potential dividing point k_i along the direction of searching for a data stream dividing point. Therefore, a speed of searching for a data stream dividing point is increased. When ~~in a preset specification,~~ a probability that at least a part of data in the window W_x [k- A_x, k+ B_x] that is determined according to the preset rule meets the preset condition C_x is 1/2, that is, a probability of executing skipping is 1/2, at most $\| B_{11} \| + \| A_{11} \| = 189$ bytes can be skipped each time.

[0164] In this implementation manner, a preset rule is that: 11 windows W_x [k- A_x, k+ B_x] are determined for a potential dividing point k and at least a part of data in the window W_x [k- A_x, k+ B_x] meets a preset condition C_x, where a probability that at least a part of data in W_x [k- A_x, k+ B_x] meets the preset condition C_x is 1/2, where x indicates consecutive natural numbers from 1 to 11 separately, and A_x and B_x are integers. A_1=169, B_1=0; A_2=170, B_2=-1; A_3=171, B_3=-2; A_4=172, B_4=-3; A_5=173, B_5=-4; A_6=174, B_6=-5; A_7=175, B_7=-6; A_8=176, B_8=-7; A_9=177, B_9=-8; A_{10}=178, B_{10}=-9; A_{11}=179, B_{11}=-10. $C_1 = C_2 = C_3 = C_4 = C_5 = C_6 = C_7 = C_8 = C_9 = C_{10} = C_{11}$. That is, 11 windows are selected for the potential dividing point k, and the 11 windows are consecutive; P(n) can be calculated by using the two factors, that is, the quantity of windows and the probability that at least a part of data in the window W_x [p_x - A_x, p_x + B_x] meets the preset condition C_x. A manner of selecting the 11 windows and the determining that at least a part of data

in each window of the 11 windows meets the preset condition C_x follow the rule preset on the deduplication server 103, and therefore, whether the potential dividing point k is a data stream dividing point depends on whether it exists that at least a part of data in each window of 11 consecutive windows meets the preset condition C_x. A gap between two bytes is referred to as one point. P(n) represents a probability that 11 consecutive windows meeting a condition do not exist among n consecutive windows, that is, a probability that no data stream dividing point exists. After a minimum chunk size of 4 KB is skipped from a file header/previous dividing point, a 4086th point is found by going back by 10 bytes in a direction opposite to the direction of searching for a data stream dividing point, and no data stream dividing point exists at the point; therefore, P(4086)=1, and P(4087)=1, ..., P(4095)=1, and so on. At an 4096th point, that is, a point which is used to obtain the minimum chunk, with a probability of (1/2)^11, at least a part of data in each window of the 11 windows meets the preset condition C_x. Hence, with a probability of (1/2)^11, a data stream dividing point exists; with a probability of 1-(1/2)^11, no data stream dividing point exists; therefore P(4096)=1-(1/2)^11.

[0165] In an nth window, there may be 12 cases of obtaining P(n) by means of recursion.



[0166] Case 1: With a probability of 1/2, at least a part of data in the nth window does not meet a preset condition; in this case, with a probability of P(n-1), 11 consecutive windows do not exist among (n-1) windows before the nth window, where at least a part of data in each window of the 11 consecutive windows meets a preset condition. Therefore, P(n) includes 1/2*P(n-1). A case in which the at least a part of data in the nth window does not meet the preset condition, and 11 consecutive windows exist among the (n-1) windows before the nth ~~point~~window, where at least a part of data in each window of the 11 consecutive windows meets the preset condition, is not related to P(n).

depends on whether it exists that at least a part of data in each window of the 24 consecutive windows meets the preset condition C_x, and calculation can be performed by using the following formulas:

P(1)=1, P(2), ..., P(23)=1, P(24)=1-(3/4)^24, and

P(n)=1/4*P(n-1)+1/4*(3/4)*P(n-2)+...+1/4*(3/4)^23*P(n-24).

[0172] After calculation, P(5*1024)=0.78, P(11*1024)=0.17, and P(12*1024)=0.13. That is, no data stream dividing point is found with a probability of 13% after a search proceeds to a point at a distance of 12 KB from a start position/previous data stream dividing point of a data stream, and forced division is performed. A density function of a data stream dividing point is obtained by using this probability, and after integration, it is obtained that on average, a data stream dividing point is found after a search proceeds to a point at a distance about 7.6 KB from the start position/previous data stream dividing point of the data stream, that is, an average chunk length is about 7.6 KB. Different from that at least a part of data in 11 consecutive windows meets a preset condition with a probability of 1/2, a conventional CDC algorithm can achieve an effect of an average chunk length being 7.6 KB only when one window meets a condition with a probability of 1/2^12.

[0173] On the basis of the search for a data stream dividing point shown in FIG. 3, in an implementation manner shown in FIG. 23, a rule is preset on a deduplication server 103, where the rule is: for a potential dividing point k, determining 11 windows $W_x[k-A_x, k+B_x]$ and a preset condition C_x corresponding to the window $W_x[k-A_x, k+B_x]$, where x indicates consecutive natural numbers from 1 to 11, and A_x and B_x are integers. A probability that at least a part of data in the window $W_x[k-A_x, k+B_x]$ meets the preset condition C_x is 1/2. A_1=171, B_1=-2; A_2=172, B_2=-3; A_3=173, B_3=-4; A_4=174, B_4=-5; A_5=175, B_5=-6; A_6=176, B_6=-7; A_7=177, B_7=-8; A_8=178, B_8=-9; A_9=179, B_9=-10; A_{10}=170, B_{10}=-1; A_{11}=169, B_{11}=0. $C_1=C_2=C_3=C_4=C_5=C_6=C_7=C_8=C_9=C_{10}=C_{11}$. k_a is a data stream dividing point, and a direction of searching for a data stream dividing point shown in FIG. 23 is from left to right. After a minimum data chunk of 4 KB is skipped from the data stream dividing point k_a, an end position of the minimum data chunk of 4 KB serves as a next potential dividing point k_i. According to the rule preset on the deduplication server 103, for the potential dividing point k_i, $W_{ix}[k_i-A_x, k_i+B_x]$ ~~$W_x[k-A_x, k+B_x]$~~ and the preset condition C_x corresponding to the window $W_{ix}[k_i-A_x, k_i+B_x]$ ~~$W_x[k-A_x, k+B_x]$~~ are determined, where x indicates consecutive natural numbers from 1 to 11. Determined 11 windows are $W_{i1}[k_i-171, k_i-2]$, $W_{i2}[k_i-172, k_i-3]$, $W_{i3}[k_i-173, k_i-4]$, $W_{i4}[k_i-174, k_i-5]$, $W_{i5}[k_i-175, k_i-6]$, $W_{i6}[k_i$

least a part of data in the window W_{i7} meets the preset condition C_7, the at least a part of data in the window W_{i8} meets the preset condition C_8, the at least a part of data in the window W_{i9} meets the preset condition C_9, the at least a part of data in the window W_{i10} meets the preset condition C_{10}, and the at least a part of data in the window W_{i11} meets the preset condition C_{11}, the current potential dividing point k_i is a data stream dividing point. When the at least a part of data in any window of the 11 windows does not meet the corresponding preset condition, for example, as shown in FIG. 30, the at least a part of data in $W_{i4}[k_i\text{-}175, k_i\text{-}6]$ does not meet the preset condition C_4, a next potential dividing point is selected. For differentiation from the potential dividing point k_i, the next potential dividing point is represented as k_j here, where k_j is located on the right of k_i, and a distance between k_j and k_i is 1 byte. As shown in FIG. 30, according to the rule preset ~~for~~ on the deduplication server 103, for the potential dividing point k_j, 11 windows are determined, and are $W_{j1}[k_j\text{-}169, k_j]$, $W_{j2}[k_j\text{-}171, k_j\text{-}2]$, $W_{j3}[k_j\text{-}173, k_j\text{-}4]$, $W_{j4}[k_j\text{-}175, k_j\text{-}6]$, $W_{j5}[k_j\text{-}177, k_j\text{-}8]$, $W_{j6}[k_j\text{-}179, k_j\text{-}10]$, $W_{j7}[k_j\text{-}181, k_j\text{-}12]$, $W_{j8}[k_j\text{-}183, k_j\text{-}14]$, $W_{j9}[k_j\text{-}185, k_j\text{-}16]$, $W_{j10}[k_j\text{-}187, k_j\text{-}18]$, and $W_{j11}[k_j\text{-}189, k_j\text{-}20]$ separately, where $C_1=C_2=C_3=C_4=C_5=C_6=C_7=C_8=C_9=C_{10}=C_{11}$. It is determined whether at least a part of data in $W_{j1}[k_j\text{-}169, k_j]$ meets the preset condition C_1, it is determined whether at least a part of data in $W_{j2}[k_j\text{-}171, k_j\text{-}2]$ meets the preset condition C_2, it is determined whether at least a part of data in $W_{j3}[k_j\text{-}173, k_j\text{-}4]$ meets the preset condition C_3, it is determined whether at least a part of data in $W_{j4}[k_j\text{-}175, k_j\text{-}6]$ meets the preset condition C_4, it is determined whether at least a part of data in $W_{j5}[k_j\text{-}177, k_j\text{-}8]$ meets the preset condition C_5, it is determined whether at least a part of data in $W_{j6}[k_j\text{-}179, k_j\text{-}10]$ meets the preset condition C_6, it is determined whether at least a part of data in $W_{j7}[k_j\text{-}181, k_j\text{-}12]$ meets the preset condition C_7, it is determined whether at least a part of data in $W_{j8}[k_j\text{-}183, k_j\text{-}14]$ meets the preset condition C_8, it is determined whether at least a part of data in $W_{j9}[k_j\text{-}185, k_j\text{-}16]$ meets the preset condition C_9, it is determined whether at least a part of data in $W_{j10}[k_j\text{-}187, k_j\text{-}18]$ meets the preset condition C_{10}, and it is determined whether at least a part of data in $W_{j11}[k_j\text{-}189, k_j\text{-}20]$ meets the preset condition C_{11}. When it is determined that the at least a part of data in the

window W_{j1} meets the preset condition C_1, the at least a part of data in the window W_{j2} meets the preset condition C_2, the at least a part of data in the window W_{j3} meets the preset condition C_3, the at least a part of data in the window W_{j4} meets the preset condition C_4, the at least a part of data in the window W_{j5} meets the preset condition C_5, the at least a part of data in the window W_{j6} meets the preset condition C_6, the at least a part of data in the window W_{j7} meets the preset condition C_7, the at least a part of data in the window W_{j8} meets the preset condition C_8, the at least a part of data in the window W_{i9} meets the preset condition C_9, the at least a part of data in the window W_{j10} meets the preset condition C_{10}, and the at least a part of data in the window W_{j11} meets the preset condition C_{11}, the current potential dividing point k_j is a data stream dividing point. When it is determined that the at least a part of data in any window of the windows W_{j1}, W_{j2}, W_{j3}, W_{j4}, W_{j5}, W_{j6}, W_{j7}, W_{j8}, W_{j9}, W_{j10}, and W_{j11} does not meet a preset condition, for example, as shown in FIG. 31, when the at least a part of data in $W_{j3}[k_j\text{-}173,\ k_j\text{-}4]$ does not meet the preset condition C_3, k_j is located on the right of k_i and N bytes are skipped from k_i along the direction of searching for a data stream dividing point, where the N bytes are not greater than $\| B_4 \| + \max_x(\| A_x \|)$. In the implementation manner shown in FIG. 28, the N bytes are not greater than 195 bytes, and in this embodiment, N=15, so as to obtain a next potential dividing point. For differentiation from the potential dividing points k_i and k_j, the next potential dividing point is represented as k_l. According to the rule preset ~~for~~ on the deduplication server 103 in the implementation manner in FIG. 29, for the potential dividing point k_l, 11 windows are determined, and are $W_{l1}[k_l\text{-}169,\ k_l]$, $W_{l2}[k_l\text{-}171,\ k_l\text{-}2]$, $W_{l3}[k_l\text{-}173,\ k_l\text{-}4]$, $W_{l4}[k_l\text{-}175,\ k_l\text{-}6]$, $W_{l5}[k_l\text{-}177,\ k_l\text{-}8]$, $W_{l6}[k_l\text{-}179,\ k_l\text{-}10]$, $W_{l7}[k_l\text{-}181,\ k_l\text{-}12]$, $W_{l8}[k_l\text{-}183,\ k_l\text{-}14]$, $W_{l9}[k_l\text{-}185,\ k_l\text{-}16]$, $W_{l10}[k_l\text{-}187,\ k_l\text{-}18]$, and $W_{l11}[k_l\text{-}189,\ k_l\text{-}20]$ separately. It is determined whether at least a part of data in $W_{l1}[k_l\text{-}169,\ k_l]$ meets the preset condition C_1, it is determined whether at least a part of data in $W_{l2}[k_l\text{-}171,\ k_l\text{-}2]$ meets the preset condition C_2, it is determined whether at least a part of data in $W_{l3}[k_l\text{-}173,\ k_l\text{-}4]$ meets the preset condition C_3, it is determined whether at least a part of data in $W_{l4}[k_l\text{-}175,\ k_l\text{-}6]$ meets the preset condition C_4, it is determined whether at least a part of data in $W_{l5}[k_l\text{-}177,\ k_l\text{-}8]$ meets the preset condition C_5, it is determined whether at least a part of data in $W_{l6}[k_l\text{-}179,\ k_l\text{-}10]$ meets the preset condition

C_6, it is determined whether at least a part of data in $W_{l7}[k_l\text{-}181,\ k_l\text{-}12]$ meets the preset condition C_7, it is determined whether at least a part of data in $W_{l8}[k_l\text{-}183,\ k_l\text{-}14]$ meets the preset condition C_8, it is determined whether at least a part of data in $W_{l9}[k_l\text{-}185,\ k_l\text{-}16]$ meets the preset condition C_9, it is determined whether at least a part of data in $W_{l10}[k_l\text{-}187,\ k_l\text{-}18]$ meets the preset condition C_{10}, and it is determined whether at least a part of data in $W_{l11}[k_l\text{-}189,\ k_l\text{-}20]$ meets the preset condition C_{11}. When it is determined that the at least a part of data in the window W_{l1} meets the preset condition C_1, the at least a part of data in the window W_{l2} meets the preset condition C_2, the at least a part of data in the window W_{l3} meets the preset condition C_3, the at least a part of data in the window W_{l4} meets the preset condition C_4, the at least a part of data in the window W_{l5} meets the preset condition C_5, the at least a part of data in the window W_{l6} meets the preset condition C_6, the at least a part of data in the window W_{l7} meets the preset condition C_7, the at least a part of data in the window W_{l8} meets the preset condition C_8, the at least a part of data in the window W_{l9} meets the preset condition C_9, the at least a part of data in the window W_{l10} meets the preset condition C_{10}, and the at least a part of data in the window W_{l11} meets the preset condition C_{11}, the current potential dividing point k_l is a data stream dividing point. When the at least a part of data in any window of the windows W_{l1}, W_{l2}, W_{l3}, W_{l4}, W_{l5}, W_{l6}, W_{l7}, W_{l8}, W_{l9}, W_{l10}, and W_{l11} does not meet a preset condition, a next potential dividing point is selected. For differentiation from the potential dividing points k_i, k_j, and k_l, the next potential dividing point is represented as k_m, where k_m is located on the right of k_l, and a distance between k_m and k_l is 1 byte. According to the rule preset on ~~for~~ the deduplication server 103 in the embodiment shown in FIG. 29, 11 windows determined for the potential dividing point k_m are $W_{m1}[k_m\text{-}169,\ k_m]$, $W_{m2}[k_m\text{-}171,\ k_m\text{-}2]$, $W_{m3}[k_m\text{-}173,\ k_m\text{-}4]$, $W_{m4}[k_m\text{-}175,\ k_m\text{-}6]$, $W_{m5}[k_m\text{-}177,\ k_m\text{-}8]$, $W_{m6}[k_m\text{-}179,\ k_m\text{-}10]$, $W_{m7}[k_m\text{-}181,\ k_m\text{-}12]$, $W_{m8}[k_m\text{-}183,\ k_m\text{-}14]$, $W_{m9}[k_m\text{-}185,\ k_m\text{-}16]$, $W_{m10}[k_m\text{-}187,\ k_m\text{-}18]$, and $W_{m11}[k_m\text{-}189,\ k_m\text{-}20]$ separately. It is determined whether at least a part of data in $W_{m1}[k_m\text{-}169,\ k_m]$ meets the preset condition C_1, it is determined whether at least a part of data in $W_{m2}[k_m\text{-}171,\ k_m\text{-}2]$ meets the preset condition C_2, it is determined whether at least a part of data in $W_{m3}[k_m\text{-}173,\ k_m\text{-}4]$ meets the preset condition C_3, it is determined whether at least a part of data in $W_{m4}[k_m\text{-}175,\ k_m\text{-}6]$ meets the

condition; therefore the matrix R is used: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{255,1} & h_{255,2} & \cdots & h_{255,8} \end{pmatrix}$, and random numbers of an mth row of the matrix V_b and the mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{bm} = V_{bm,1} * h_{m,1} + V_{bm,2} * h_{m,2} + ... + V_{bm,8} * h_{m,8}$. S_{b1}, S_{b2}, ..., and S_{b255} are obtained according to the method, and a quantity K of values meeting a specific condition (being greater than 0 is used as an example here) among S_{b1}, S_{b2}, ..., and S_{b255} is counted. Because the matrix R follows normal distribution, S_{bm} still follows normal distribution as the matrix R does. According to the probability theory, the probability that a random number in normal distribution is greater than 0 is 1/2; a probability that each value among S_{b1}, S_{b2}, ..., and S_{b255} is greater than 0 is 1/2, and therefore, K meets binomial distribution: $P(k=n) = C_{255}^{n}(\frac{1}{2})^{n}(\frac{1}{2})^{255-n} = C_{255}^{n}(\frac{1}{2})^{255}$. According to a counting result, it is determined whether the quantity K of values greater than 0 among S_{b1}, S_{b2}, ..., and S_{b255} is an even number; the probability that a random number in binomial distribution is an even number is 1/2, and therefore, K meets a condition with a probability of 1/2. When K is an even number, it indicates that the at least a part of data in $W_{i2}[k_i\text{-}170, k_i\text{-}1]$ meets the preset condition C_2. When K is an odd number, it indicates that the at least a part of data in $W_{i2}[k_i\text{-}170, k_i\text{-}1]$ does not meet the preset condition C_2. C_2 here refers to that the quantity K, which is obtained according to the foregoing manner, of values greater than 0 among S_{b1}, S_{b2}, ..., and S_{b255} is an even number. In the implementation manner shown in FIG. 21, the at least a part of data in $W_{i2}[k_i\text{-}170, k_i\text{-}1]$ meets the preset condition C_2.

[0178] Therefore, as shown in FIG. 32, "▨" represents 1 byte selected when it is determined whether at least a part of data in the window $W_{i3}[k_i\text{-}171, k_i\text{-}2]$ meets a preset condition C_3, and there are 42 bytes between two adjacent selected bytes. Selected 5 bytes of data are used repeatedly 51 times to obtain 255 bytes in total, so as to increase randomness. Then, the method for determining whether at least a part of data in the windows $W_{i1}[k_i\text{-}169, k_i]$ and $W_{i2}[k_i\text{-}170, k_i\text{-}1]$ meets a preset condition is used to determine whether at least a part of data in $W_{i3}[k_i\text{-}171, k_i\text{-}2]$ meets the preset condition C_3. In the implementation manner shown in FIG. 21, the at least a part of data

in $W_{i3}[k_i\text{-}171,\ k_i\text{-}2]$ meets the preset condition. As shown in FIG. 32, "▨" represents 1 byte selected when it is determined whether at least a part of data in the window $W_{i4}[k_i\text{-}172,\ k_i\text{-}3]$ meets a preset condition C_4, and there are 42 bytes between two adjacent selected bytes. Selected 5 bytes of data are used repeatedly 51 times to obtain 255 bytes in total, so as to increase randomness. Then, the method for determining whether at least a part of data in the windows $W_{i1}[k_i\text{-}169,\ k_i]$, $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$, and $W_{i3}[k_i\text{-}171,\ k_i\text{-}2]$ meets a preset condition is used to determine whether the at least a part of data in $W_{i4}[k_i\text{-}172,\ k_i\text{-}3]$ meets the preset condition C_4. In the implementation manner shown in FIG. 21, the at least a part of data in $W_{i4}[k_i\text{-}172,\ k_i\text{-}3]$ meets the preset condition C_4. As shown in FIG. 32, "▦" represents 1 byte selected when it is determined whether at least a part of data in the window $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ meets a preset condition C_5, and there are 42 bytes between two adjacent selected bytes. Selected 5 bytes of data are used repeatedly 51 times to obtain 255 bytes in total, so as to increase randomness. Then, the method for determining whether at least a part of data in the windows $W_{i1}[k_i\text{-}169,\ k_i]$, $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$, $W_{i3}[k_i\text{-}171,\ k_i\text{-}2]$, and $W_{i4}[k_i\text{-}172,\ k_i\text{-}3]$ meets a preset condition is used to determine whether the at least a part of data in $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ meets the preset condition C_5. In the implementation manner shown in FIG. 21, the at least a part of data in $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ does not meet the preset condition C_5.

[0179] When the at least a part of data in $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ does not meet the preset condition C_5, 7 bytes are skipped from a point p_{i5} along a direction of searching for a data stream dividing point, and a next potential dividing point k_j is obtained at an end position of a 7th byte. As shown in FIG. 22, according to the rule preset on ~~for~~ the deduplication server 103, a window $W_{j1}[k_j\text{-}169,\ k_j]$ is determined for the potential dividing point k_j. A manner of determining whether at least a part of data in the window $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1 is the same as a manner of determining whether at least a part of data in the window $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1. Therefore, as shown in FIG. 33, W_{j1} represents a window, and 5 bytes are selected to determine whether the at least a part of data in the window meets the preset condition C_1. In FIG. 33, "■" represents 1 selected byte, and there are 42 bytes between two adjacent selected bytes. The selected 5 bytes of data are used repeatedly 51 times to obtain 255 bytes in total,

determined whether at least a part of data in $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1. As shown in FIG. 32, W_{i1} represents the window $W_{i1}[k_i\text{-}169,\ k_i]$, and 5 bytes are selected to determine whether the at least a part of data in $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1. In FIG. 32, "■" represents 1 selected byte, and there are 42 bytes between two adjacent selected bytes "■". In one implementation manner, a hash function is used to calculate the selected 5 bytes, and values obtained by means of calculation by using the hash function are in fixed and even distribution. If a value obtained by means of calculation by using the hash function is an even number, it is determined that the at least a part of data in $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1. That is, C_1 represents that a value obtained according to the foregoing manner and by means of calculation by using the hash function is an even number. Therefore, a probability that ~~whether~~ at least a part of data in $W_{i1}[k_i\text{-}169,\ k_i]$ meets a preset condition is 1/2. In the implementation manner shown in FIG. 21, it is determined, by using a hash function, whether at least a part of data in $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$ meets a preset condition C_2, whether at least a part of data in $W_{i3}[k_i\text{-}171,\ k_i\text{-}2]$ meets a preset condition C_3, whether at least a part of data in $W_{i4}[k_i\text{-}172,\ k_i\text{-}3]$ meets a preset condition C_4, and whether at least a part of data in $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ meets a preset condition C_5. For specific implementation, reference may be made to the description of the manner of determining, by using a hash function, whether at least a part of data in $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1 in the implementation manner shown in FIG. 21, which is not described herein again.

[0182] When the at least a part of data in $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ does not meet the preset condition C_5, 7 bytes are skipped from the potential dividing point k_i along a direction of searching for a data stream dividing point, and a current potential dividing point k_j is obtained at an end position of a 7th byte. As shown in FIG. 22, according to the rule preset on ~~for~~ the deduplication server 103, a window $W_{j1}[k_j\text{-}169,\ k_j]$ is determined for the potential dividing point k_j. A manner of determining whether at least a part of data in the window $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1 is the same as a manner of determining whether at least a part of data in the window $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1. Therefore, as shown in FIG. 33, W_{j1} represents the window $W_{j1}[k_j\text{-}169,\ k_j]$, and 5 bytes are selected to determine whether the at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1. In FIG. 33, "■" represents 1 selected

byte, and there are 42 bytes between two adjacent selected bytes "■". The 5 bytes selected from the window $W_{j1}[k_j\text{-}169,\ k_j]$ are calculated by using a hash function. If an obtained value is an even number, the at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1. In FIG. 33, a manner of determining whether at least a part of data in $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$ meets the preset condition C_2 is the same as a manner of determining whether at least a part of data in $W_{j2}[k_j\text{-}170,\ k_j\text{-}1]$ meets the preset condition C_2. Therefore, as shown in FIG. 33, " " represents 1 byte selected when it is determined whether the at least a part of data in the window $W_{j2}[k_j\text{-}170,\ k_j\text{-}1]$ meets the preset condition C_2, and there are 42 bytes between two adjacent selected bytes " ". Selected 5 bytes are calculated by using a hash function. If an obtained value is an even number, the at least a part of data in $W_{j2}[k_j\text{-}170,\ k_j\text{-}1]$ meets the preset condition C_2. In FIG. 33, a manner of determining whether at least a part of data in $W_{i3}[k_i\text{-}171,\ k_i\text{-}2]$ meets the preset condition C_3 is the same as a manner of determining whether at least a part of data in $W_{j3}[k_j\text{-}171,\ k_j\text{-}2]$ meets the preset condition C_3. Therefore, as shown in FIG. 33, " " represents 1 byte selected when it is determined whether the at least a part of data in the window $W_{j3}[k_j\text{-}171,\ k_j\text{-}2]$ meets the preset condition C_3, and there are 42 bytes between two adjacent selected bytes " ". Selected 5 bytes are calculated by using a hash function. If an obtained value is an even number, the at least a part of data in $W_{j3}[k_j\text{-}171,\ k_j\text{-}2]$ meets the preset condition C_3. In FIG. 33, a manner of determining whether at least a part of data in $W_{j4}[k_j\text{-}172,\ k_j\text{-}3]$ meets the preset condition C_4 is the same as ~~and~~ a manner of determining whether at least a part of data in the window $W_{i4}[k_i\text{-}172,\ k_i\text{-}3]$ meets the preset condition C_4. Therefore, as shown in FIG. 33, " " represents 1 byte selected when it is determined whether the at least a part of data in the window $W_{j4}[k_j\text{-}172,\ k_j\text{-}3]$ meets the preset condition C_4, and there are 42 bytes between two adjacent selected bytes " ". Selected 5 bytes are calculated by using a hash function. If an obtained value is an even number, the at least a part of data in $W_{j4}[k_j\text{-}172,\ k_j\text{-}3]$ meets the preset condition C_4. According to the foregoing method, it is determined whether at least a part of data in $W_{j5}[k_j\text{-}173,\ k_j\text{-}4]$ meets the preset condition C_5, it is determined whether at least a part of data in $W_{j6}[k_j\text{-}174,\ k_j\text{-}5]$ meets a preset condition C_6, it is determined whether at least a part of data in $W_{j7}[k_j\text{-}175,\ k_j\text{-}6]$ meets a

b_4 is located in a 4th column, a value corresponding to $h_{26,4}$ is searched for. The matrix R is searched for a corresponding value according to the value of b_5 and a column in which b_5 is located. For example, if b_5=90, and b_5 is located in a 5th column, a value corresponding to $h_{90,5}$ is searched for. $S_2 = h_{66,1} + h_{48,2} + h_{99,3} + h_{26,4} + h_{90,5}$, and because the matrix R follows binomial distribution, S_2 also follows binomial distribution. When S_2 is an even number, the at least a part of data in $W_{i2}[k_i\text{-}170, k_i\text{-}1]$ meets the preset condition C_2; when S_2 is an odd number, the at least a part of data in $W_{i2}[k_i\text{-}170, k_i\text{-}1]$ does not meet the preset condition C_2. A probability that S_2 is an even number is 1/2. In the embodiment shown in FIG. 21, the at least a part of data in $W_{i2}[k_i\text{-}170, k_i\text{-}1]$ meets the preset condition C_2. By using the same rule, separately, it is determined whether at least a part of data in $W_{i3}[k_i\text{-}171, k_i\text{-}2]$ meets a preset condition C_3, it is determined whether at least a part of data in $W_{i4}[k_i\text{-}172, k_i\text{-}3]$ meets a preset condition C_4, it is determined whether at least a part of data in $W_{i5}[k_i\text{-}173, k_i\text{-}4]$ meets a preset condition C_5, it is determined whether at least a part of data in $W_{i6}[k_i\text{-}174, k_i\text{-}5]$ meets a preset condition C_6, it is determined whether at least a part of data in $W_{i7}[k_i\text{-}175, k_i\text{-}6]$ meets a preset condition C_7, it is determined whether at least a part of data in $W_{i8}[k_i\text{-}176, k_i\text{-}7]$ meets a preset condition C_8, it is determined whether at least a part of data in $W_{i9}[k_i\text{-}177, k_i\text{-}8]$ meets a preset condition C_9, it is determined whether at least a part of data in $W_{i10}[k_i\text{-}178, k_i\text{-}9]$ meets a preset condition C_{10}, and it is determined whether at least a part of data in $W_{i11}[k_i\text{-}179, k_i\text{-}10]$ meets a preset condition C_{11}. In the implementation manner shown in FIG. 21, the at least a part of data in $W_{i5}[k_i\text{-}173, k_i\text{-}4]$ does not meet the preset condition C_5, 7 bytes are skipped from the potential dividing point k_i along a direction of searching for a data stream dividing point, and a current potential dividing point k_j is obtained at an end position of a 7th byte. As shown in FIG. 22, according to the rule preset on ~~for~~ the deduplication server 103, a window $W_{j1}[k_j\text{-}169, k_j]$ is determined for the potential dividing point k_j. A manner of determining whether at least a part of data in the window $W_{j1}[k_j\text{-}169, k_j]$ meets the preset condition C_1 is the same as a manner of determining whether at least a part of data in the window $W_{i1}[k_i\text{-}169, k_i]$ meets the preset condition C_1. Therefore, as shown in FIG. 33, W_{j1} represents the window $W_{j1}[k_j\text{-}169, k_j]$, and it is determined whether the at least a part of data in $W_{j1}[k_j\text{-}169, k_j]$ meets the preset condition C_1. In FIG. 33, each of bytes "■"

$S_{2h}=h_{66,1}+h_{48,2}+h_{99,3}+h_{26,4}+h_{90,5}$, and because the matrix R follows binomial distribution, S_{2h} also follows binomial distribution. $S_{2g}=g_{66,1}+g_{48,2}+g_{99,3}+g_{26,4}+g_{90,5}$, and because the matrix G follows binomial distribution, S_{2g} also follows binomial distribution. When one of S_{2h} and S_{2g} is an even number, the at least a part of data in $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$ meets the preset condition C_2; when both S_{2h} and S_{2g} are odd numbers, the at least a part of data in $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$ does not meet the preset condition C_2. A probability that one of S_{2h} and S_{2g} is an even number is 3/4. In the embodiment shown in FIG. 21, the at least a part of data in $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$ meets the preset condition C_2. By using the same rule, separately, it is determined whether at least a part of data in $W_{i3}[k_i\text{-}171,\ k_i\text{-}2]$ meets a preset condition C_3, it is determined whether at least a part of data in $W_{i4}[k_i\text{-}172,\ k_i\text{-}3]$ meets a preset condition C_4, it is determined whether at least a part of data in $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ meets a preset condition C_5, it is determined whether at least a part of data in $W_{i6}[k_i\text{-}174,\ k_i\text{-}5]$ meets a preset condition C_6, it is determined whether at least a part of data in $W_{i7}[k_i\text{-}175,\ k_i\text{-}6]$ meets a preset condition C_7, it is determined whether at least a part of data in $W_{i8}[k_i\text{-}176,\ k_i\text{-}7]$ meets a preset condition C_8, it is determined whether at least a part of data in $W_{i9}[k_i\text{-}177,\ k_i\text{-}8]$ meets a preset condition C_9, it is determined whether at least a part of data in $W_{i10}[k_i\text{-}178,\ k_i\text{-}9]$ meets a preset condition C_{10}, and it is determined whether at least a part of data in $W_{i11}[k_i\text{-}179,\ k_i\text{-}10]$ meets a preset condition C_{11}. In the implementation manner shown in FIG. 21, the at least a part of data in $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ does not meet the preset condition C_5, 7 bytes are skipped from the potential dividing point k_i along a direction of searching for a data stream dividing point, and a current potential dividing point k_j is obtained at an end position of a 7th byte. As shown in FIG. 22, according to the rule preset on ~~for~~ the deduplication server 103, a window $W_{j1}[k_j\text{-}169,\ k_j]$ is determined for the potential dividing point k_j. A manner of determining whether at least a part of data in the window $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1 is the same as a manner of determining whether at least a part of data in the window $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1. Therefore, as shown in FIG. 33, W_{j1} represents the window $W_{j1}[k_j\text{-}169,\ k_j]$, and it is determined whether the at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1. In FIG. 33, each of bytes "■" whose sequence numbers are 169, 127, 85, 43, and 1 represents 1 selected byte, and there are 42 bytes

an even number is 3/4.

[0191] A manner of determining whether at least a part of data in $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$ meets the preset condition C_2 is the same as a manner of determining whether at least a part of data in $W_{j2}[k_j\text{-}170,\ k_j\text{-}1]$ meets the preset condition C_2. Therefore, as shown in FIG. 33, "▨" represents 1 byte selected when it is determined whether the at least a part of data in the window $W_{j2}[k_j\text{-}170,\ k_j\text{-}1]$ meets the preset condition C_2, and there are 42 bytes between two adjacent selected bytes. In FIG. 33, selected bytes are represented as sequence numbers 170, 128, 86, 44, and 2 separately, and there are 42 bytes between two adjacent selected bytes. The bytes "▨" whose sequence numbers are 170, 128, 86, 44, and 2 are converted into decimal values that are represented as b_1', b_2', b_3', b_4', and b_5' respectively. Because 1 byte is formed by 8 bits, each byte "▨" serves as a value, and any b_s' of b_1', b_2', b_3', b_4', and b_5' meets $0 \le b_s' \le 255$. b_1', b_2', b_3', b_4', and b_5' form a 1*5 matrix. The matrices R and G the same as those used when it is determined whether the at least a part of data in the window $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$ ~~$W_{j2}[k_j\text{-}170,\ k_j\text{-}1]$~~ meets the preset condition C_2 are used. According to the value of b_1' and a column in which b_1' is located, for example, b_1'=210, and b_1' is located in a 1st column, the matrix R is searched for a value corresponding to $h_{210,1}$, and the matrix G is searched for a value corresponding to $g_{210,1}$. According to the value of b_2' and a column in which b_2' is located, for example, b_2'=156, and b_2' is located in a 2nd column, the matrix R is searched for a value corresponding to $h_{156,2}$, and the matrix G is searched for a value corresponding to $g_{156,2}$. According to the value of b_3' and a column in which b_3' is located, for example, b_3'=144, and b_3' is located in a 3rd column, the matrix R is searched for a value corresponding to $h_{144,3}$, and the matrix G is searched for a value corresponding to $g_{144,3}$. According to the value of b_4' and a column in which b_4' is located, for example, b_4'=60, and b_4' is located in a 4th column, the matrix R is searched for a value corresponding to $h_{60,4}$, and the matrix G is searched for a value corresponding to $g_{60,4}$. According to the value of b_5' and a column in which b_5' is located, for example, b_5'=90, and b_5' is located in a 5th column, the matrix R is searched for a value corresponding to $h_{90,5}$, and the matrix G is searched for a value corresponding to $g_{90,5}$. $S_{2h}' = h_{210,1} + h_{156,2} + h_{144,3} + h_{60,4} + h_{90,5}$, and $S_{2g}' = g_{210,1} + g_{156,2} + g_{144,3} + g_{60,4} + g_{90,5}$. When one of S_{2h}' and S_{2g}' is an even number, the at least a part of data in $W_{j2}[k_j$

whether at least a part of data in $W_{i7}[k_i\text{-}175,\ k_i\text{-}6]$ meets a preset condition C_7, it is determined whether at least a part of data in $W_{i8}[k_i\text{-}176,\ k_i\text{-}7]$ meets a preset condition C_8, it is determined whether at least a part of data in $W_{i9}[k_i\text{-}177,\ k_i\text{-}8]$ meets a preset condition C_9, it is determined whether at least a part of data in $W_{i10}[k_i\text{-}178,\ k_i\text{-}9]$ meets a preset condition C_{10}, and it is determined whether at least a part of data in $W_{i11}[k_i\text{-}179,\ k_i\text{-}10]$ meets a preset condition C_{11}. In the implementation manner shown in FIG. 21, the at least a part of data in $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ does not meet the preset condition C_5, 7 bytes are skipped from the potential dividing point k_i along a direction of searching for a data stream dividing point, and a current potential dividing point k_j is obtained at an end position of a 7th byte. As shown in FIG. 22, according to the rule preset on ~~for~~ the deduplication server 103, a window $W_{j1}[k_j\text{-}169,\ k_j]$ is determined for the potential dividing point k_j. A manner of determining whether at least a part of data in the window $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1 is the same as the manner of determining whether at least a part of data in the window $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1. Therefore, as shown in FIG. 33, W_{j1} represents the window $W_{j1}[k_j\text{-}169,\ k_j]$, and 5 bytes are selected to determine whether at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1. In FIG. 33, each of bytes "■" whose sequence numbers are 169, 127, 85, 43, and 1 in FIG. 33 represents 1 selected byte, and there are 42 bytes between two adjacent selected bytes. The bytes "■" whose sequence numbers are 169, 127, 85, 43, and 1 are regarded as 40 sequential bits, which are represented as a_1', a_2', a_3', a_4', ..., and a_{40}' separately. For any a_t' of a_1', a_2', a_3', a_4', ..., and a_{40}', when $a_t'=0$, $V_{at}'=-1$, and when $a_t'=1$, $V_{at}'=1$. According to a correspondence between a_t' and V_{at}', V_{a1}', V_{a2}', V_{a3}', V_{a4}', ..., and V_{a40}' are generated. The manner of determining whether at least a part of data in the window $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1 is the same as the manner of determining whether at least a part of data in the window $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1. Therefore, the same random numbers are used: h_1, h_2, h_3, h_4, ..., and h_{40}. $S_a'=V_{a1}'*h_1+V_{a2}'*h_2+V_{a3}'*h_3+V_{a4}'*h_4+...+V_{a40}'*h_{40}$. Because h_1, h_2, h_3, h_4, ..., and h_{40} follow normal distribution, S_a' also follows normal distribution. When S_a' is a positive number, the at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1; when S_a' is a negative number or 0, the at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ does not meet the preset

[0197] When the at least a part of data in $W_{i5}[k_i\text{-}173,\ k_i\text{-}4]$ does not meet the preset condition C_5, 7 bytes are skipped from the potential dividing point k_i along a direction of searching for a data stream dividing point, and a current potential dividing point k_j is obtained at an end position of a 7th byte. As shown in FIG. 22, according to the rule preset on ~~for~~ the deduplication server 103, a window $W_{j1}[k_j\text{-}169,\ k_j]$ is determined for the potential dividing point k_j. A manner of determining whether at least a part of data in the window $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1 is the same as a manner of determining whether at least a part of data in the window $W_{i1}[k_i\text{-}169,\ k_i]$ meets the preset condition C_1. Therefore, the same correspondence R between each decimal number in 0-(2^40-1) and a designated value is used. As shown in FIG. 33, W_{j1} represents the window, and 5 bytes are selected to determine whether the at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1. In FIG. 33, "■" represents 1 selected byte, and there are 42 bytes between two adjacent selected bytes "■". Each of bytes "■" whose sequence numbers 169, 127, 85, 43, and 1 is converted into 1 decimal number, and R is searched for a designated value corresponding to the decimal number. If the designated value is an even number, the at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1; if the designated value is an odd number, the at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ does not meet the preset condition C_1. Because a probability that a random number in even distribution is an even number is 1/2, a probability that the at least a part of data in $W_{j1}[k_j\text{-}169,\ k_j]$ meets the preset condition C_1 is 1/2. Similarly, a manner of determining whether at least a part of data in $W_{i2}[k_i\text{-}170,\ k_i\text{-}1]$ meets the preset condition C_2 is the same as a manner of determining whether at least a part of data in $W_{j2}[k_j\text{-}170,\ k_j\text{-}1]$ meets the preset condition C_2, and a manner of determining whether at least a part of data in $W_{i3}[k_i\text{-}171,\ k_i\text{-}2]$ meets the preset condition C_3 is the same as a manner of determining whether at least a part of data in $W_{j3}[k_j\text{-}171,\ k_j\text{-}2]$ meets the preset condition C_3. Similarly, it is determined whether at least a part of data in $W_{j4}[k_j\text{-}172,\ k_j\text{-}3]$ meets the preset condition C_4, it is determined whether at least a part of data in $W_{j5}[k_j\text{-}173,\ k_j\text{-}4]$ meets the preset condition C_5, it is determined whether at least a part of data in $W_{j6}[k_j\text{-}174,\ k_j\text{-}5]$ meets the preset condition C_6, it is determined whether at least a part of data in $W_{j7}[k_j\text{-}175,\ k_j\text{-}6]$ meets the preset condition C_7, it is determined whether at least a part of data in $W_{j8}[k_j\text{-}176,\ k_j$

current potential dividing point k_i meets the preset condition C_x, select the current potential dividing point k_i as a data stream dividing point.

[0201] Further, the rule further includes that at least two windows $W_{ie}[k_i - A_e,\ k_i + B_e]$ and $W_{if}[k_i - A_f,\ k_i + B_f]$ meet conditions $|A_e + B_e| = |A_f + B_f|$ and $C_e = C_f$. Further, the rule further includes that A_e and A_f are positive integers. Further, the rule further includes $A_e - 1 = A_f$ and $B_e + 1 = B_f$.

[0202] Further, the judging and processing unit 1902 is specifically configured to determine, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z. Still further, the judging and processing unit 1902 specifically determines, by using a hash function, whether the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z.

[0203] Further, the judging and processing unit 1902 is configured to: when the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ does not meet the preset condition C_z, skip the N minimum units U for searching for a data stream dividing point from the current potential dividing point k_i along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and the determining unit 1901 performs step (a) for the new potential dividing point, where according to the rule, a left boundary of a window $W_{ic}[k_i - A_c,\ k_i + B_c]$ that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ or a left boundary of a window $W_{ic}[k_i - A_c,\ k_i + B_c]$ that is determined for the new potential dividing point falls within a range of the window $W_{iz}[k_i - A_z,\ k_i + B_z]$, where the window $W_{ic}[k_i - A_c,\ k_i + B_c]$ determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the rule.

[0204] Further, that the judging and processing unit 1902 determines, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z specifically includes:

selecting F bytes in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, where $F \geq 1$, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits

as the above, which is not described again.

[0206] In addition, according to the embodiments of the present invention shown in FIG. 20 to FIG. 33, a rule is preset on the deduplication server 103, where the rule is: determining M windows W_x[k-A_x, k+B_x] for a potential dividing point k and a preset condition C_x corresponding to the window W_x[k-A_x, k+B_x], where x indicates consecutive natural numbers from 1 to M separately, and M≥2. In the preset rule, A_1, A_2, A_3, ..., and A_m may be not all equal, B_1, B_2, B_3, ..., and B_m may be not all equal, and C_1, C_2, C_3, ..., and C_M may also be not all the same. In the implementation manner shown in FIG. 21, for W_{i1}[k_i-169, k_i], W_{i2}[k_i-170, k_i-1], W_{i3}[k_i-171, k_i-2], W_{i4}[k_i-172, k_i-3], W_{i5}[k_i-173, k_i-4], W_{i6}[k_i-174, k_i-5], W_{i7}[k_i-175, k_i-6], W_{i8}[k_i-176, k_i-7], W_{i9}[k_i-177, k_i-8], W_{i10}[k_i-178, k_i-9], and W_{i11}[k_i-179, k_i-10], the windows are the same in size, that is, all the windows have a size of 169 bytes, and meanwhile manners of determining whether at least a part of data in a window meets a preset condition are also the same. For details, refer to the foregoing description of determining whether the at least a part of data in W_{i1}[k_i-169, k_i] meets the preset condition C_1. However, in the implementation manner shown in FIG. 11, the windows W_{i1}[k_i-169, k_i], W_{i2}[k_i-170, k_i-1], W_{i3}[k_i-171, k_i-2], W_{i4}[k_i-172, k_i-3], W_{i5}[k_i-173, k_i-4], W_{i6}[k_i-174, k_i-5], W_{i7}[k_i-175, k_i-6], W_{i8}[k_i-176, k_i-7], W_{i9}[k_i-177, k_i-8], W_{i10}[k_i-168, k_i+1], and W_{i11}[k_i-179, k_i+3] may be different in size, and meanwhile the manners of determining whether at least a part of data in a window meets a preset condition may also be different. In all embodiments, according to the rule preset on ~~for~~ the deduplication server 103, a manner of determining whether at least a part of data in the window W_{i1} meets the preset condition C_1 is certainly the same as a manner of determining whether at least a part of data in a window W_{j1} meets the preset condition C_1, a manner of determining whether at least a part of data in W_{i2} meets the preset condition C_2 is certainly the same as a manner of determining whether at least a part of data in W_{j2} meets the preset condition C_2, ..., and a manner of determining whether at least a part of data in a window W_{iM} meets a preset condition C_M is certainly the same as a manner of determining whether at least a part of data in a window W_{jM} meets the preset condition C_M, which are not described herein again.

[0207] According to the embodiments of the present invention shown in FIG. 20 to FIG. 33, the rule is preset on the deduplication server 103, k_a, k_i, k_j, k_l, and k_m are potential dividing points obtained in the search for a dividing point along a direction of searching for a data stream

dividing point, and k_a, k_i, k_j, k_l, and k_m all follow the rule. In this embodiment of the present invention, the window W_x[k-A_x, k+B_x] represents a specific range, and data is selected in the specific range to determine whether the data meets the preset condition C_x. Specifically, in the specific range, a part of data may be selected or all data may be selected to determine whether the data meets the preset condition C_x. Reference may be made to the window W_x[k-A_x, k+B_x] for the concept of a window specifically used in this embodiment of the present invention, which is not described herein again.

[0208] In the window W_x[k-A_x, k+B_x], (k-A_x) and (k+B_x) represent two boundaries of the window W_x[k-A_x, k+B_x], where (k-A_x) represents a boundary, which is in a direction opposite to the direction of searching for a data stream dividing point relative to the potential dividing point k, of the window W_x[k-A_x, k+B_x], and (k+B_x) represents a boundary, which is in the direction of searching for a data stream dividing point relative to the potential dividing point k, of the window W_x[k-A_x, k+B_x]. Specifically, in the embodiment of the present invention, the direction of searching for a data stream dividing point shown in FIG. 20 to FIG. 33 is from left to right; therefore, (k-A_x) represents a boundary (that is, a left boundary), which is in a direction opposite to the direction of searching for a data stream dividing point relative to the potential dividing point k, of the window W_x[k-A_x, k+B_x], and (k+B_x) represents a boundary (that is, a right boundary), which is in the direction of searching for a data stream dividing point relative to the potential dividing point k, of the window W_x[k-A_x, k+B_x]. If the direction of searching for a data stream dividing point shown in FIG. 20 to FIG. 33 is from right to left, (k-A_x) represents a boundary (that is, a right boundary), which is in a direction opposite to the direction of searching for a data stream dividing point relative to the potential dividing point k, of the window W_x[k-A_x, k+B_x], and (k+B_x) represents a boundary (that is, a left boundary), which is in the direction of searching for a data stream dividing point relative to the potential dividing point k, of the window W_x[k-A_x, k+B_x].

[0209] A person of ordinary skill in the art may be aware that, in conjunction with various exemplary units and algorithm steps described in FIG. 20 to FIG. 33 of the embodiments of the present invention, a key feature in the embodiments of the present invention may be combined with other technologies and presented in a more complex form; however, the key feature of the present invention is still included. An alternative dividing point may be used in a real environment. For example, in an implementation manner, according to a rule preset ~~for~~ on a deduplication server 103, for a potential dividing point k_i, 11 windows W_x[k-A_x, k+B_x] and a preset condition C_x

CLAIMS

What is claimed is:

1. A method for searching for a data stream dividing point based on a server, comprising:

~~presetting a rule on the server, wherein a~~the rule preset on the server is: for a potential dividing point k, determining M points p_x, a window $W_x[\ p_x - A_x,\ p_x + B_x\ ]$ corresponding to the point p_x, and a preset condition C_x corresponding to the window $W_x[\ p_x - A_x,\ p_x + B_x\ ]$, wherein x indicates consecutive natural numbers from 1 to M, M≥2, and A_x and B_x are integers; and

the method comprising:

(a) determining a point p_{iz} and a window $W_{iz}[\ p_{iz} - A_z,\ p_{iz} + B_z\ ]$ corresponding to the point p_{iz} for a current potential dividing point k_i according to the rule, wherein i and z are integers, and 1≤z≤M;

(b) determining whether at least a part of data in the window $W_{iz}[\ p_{iz} - A_z,\ p_{iz} + B_z\ ]$ meets a preset condition C_z; and

when the at least a part of data in the window $W_{iz}[\ p_{iz} - A_z,\ p_{iz} + B_z\ ]$ does not meet the preset condition C_z, skipping N minimum units U for searching for a data stream dividing point from the point p_{iz} along a direction of searching for a data stream dividing point, wherein N*U is not greater than $\| B_z \| + \max_x(\| A_x \| + \| (k_i - p_{ix}) \|)$, so as to obtain a new potential dividing point, and performing step (a); and

(c) when at least a part of data in each window $W_{ix}[\ p_{ix} - A_x,\ p_{ix} + B_x\ ]$ of M windows of the current potential dividing point k_i meets the preset condition C_x, selecting the current potential dividing point k_i as a data stream dividing point.

2. The method according to claim 1, wherein the rule further comprises that at least two points p_e and p_f meet conditions $A_e = A_f$, $B_e = B_f$, and $C_e = C_f$.

3. The method according to claim 2, wherein the rule further comprises: relative to the potential dividing point k, the at least two points p_e and p_f are in a direction opposite to the direction of searching for a data stream dividing point.

4. The method according to claim 2 or 3, wherein the rule further comprises that a distance between the at least two points p_e and p_f is 1 U.

5. The method according to any one of claims 1 to 4, wherein the determining whether at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ meets a preset condition C_z specifically comprises:

determining, by using a random function, whether the at least a part of data in the window W_{iz} $[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ meets the preset condition C_z.

6. The method according to claim 5, wherein the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ meets the preset condition C_z is specifically: determining, by using a hash function, whether the at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ meets the preset condition C_z.

7. The method according to any one of claims 1 to 6, wherein when the at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the point p_{iz} along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[\ p_{ic}-A_c,\ p_{ic}+B_c\ ]$ corresponding to a point p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ or a left boundary of a window $W_{ic}[\ p_{ic}-A_c,\ p_{ic}+B_c\ ]$ corresponding to a point p_{ic} that is determined for the new potential dividing point falls within a range of the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$, wherein the point p_{ic} determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

8. The method according to claim 5, wherein the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ meets the preset condition C_z specifically comprises:

selecting F bytes in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, wherein F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits

corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, wherein when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, wherein $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + \ldots + V_{am,8} * h_{m,8}$, S_{a1}, S_{a2}, ..., and S_{aF*H} are obtained in a same way, a quantity K of values greater than 0 among S_{a1}, S_{a2}, ..., and S_{aF*H} is counted, and when K is an even number, the at least a part of data in the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ meets the preset condition C_z.

9. A method for searching for a data stream dividing point based on a server, comprising:

a rule preset on the server~~presetting a rule on the server, wherein the rule~~ is: for a potential dividing point k, determining M windows $W_x[k - A_x, k + B_x]$ and a preset condition C_x corresponding to the window $W_x[k - A_x, k + B_x]$, wherein x indicates consecutive natural numbers from 1 to M, M≥2, and A_x and B_x are integers; and

the method comprising:

(a) determining a corresponding window $W_{iz}[k_i - A_z, k_i + B_z]$ for a current potential dividing point k_i according to the rule, wherein i and z are integers, and 1≤z≤M;

(b) determining whether at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets a preset condition C_z; and

B_c] that is determined for the new potential dividing point coincides with a right boundary of the window W_{iz} [k_i - A_z , k_i + B_z] or a left boundary of a window W_{ic} [k_i - A_c , k_i + B_c] that is determined for the new potential dividing point falls within a range of the window W_{iz} [k_i - A_z , k_i + B_z], wherein the window W_{ic} [k_i - A_c , k_i + B_c] determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to ~~a~~ the ~~data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the rule.

16. The method according to claim 13, wherein the determining, by using a random function, whether the at least a part of data in the window W_{iz} [k_i - A_z , k_i + B_z] meets the preset condition C_z specifically comprises:

selecting F bytes in the window W_{iz} [k_i - A_z , k_i + B_z], and using the F bytes repeatedly H times to obtain F*H bytes in total, wherein F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, wherein when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, wherein $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + \ldots + V_{am,8} * h_{m,8}$, S_{a1} , S_{a2} , ..., and S_{aF*H} are obtained in a same way, a quantity K of values greater than 0 among S_{a1} , S_{a2} , ..., and S_{aF*H} is counted,

and when K is an even number, the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z.

17. A server for searching for a data stream dividing point, wherein the server comprises a central processing unit and a primary memory, the central processing unit communicates with the primary memory, and a rule preset on the server~~a rule is preset on the server, wherein the rule~~ is: for a potential dividing point k, determining M points p_x, a window $W_x[p_x - A_x, p_x + B_x]$ corresponding to the point p_x, and a preset condition C_x corresponding to the window $W_x[p_x - A_x, p_x + B_x]$, wherein x indicates consecutive natural numbers from 1 to M, M≥2, and A_x and B_x are integers; and

the main memoryis configured to store an executable instruction, and the central processing unit executes the executable instruction to perform the following steps:

(a) determining a point p_{iz} and a window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ corresponding to the point p_{iz} for a current potential dividing point k_i according to the rule, wherein i and z are integers, and 1≤z≤M;

(b) determining whether at least a part of data in the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ meets a preset condition C_z; and

when the at least a part of data in the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ does not meet the preset condition C_z, skipping N minimum units U for searching for a data stream dividing point from the point p_{iz} along a direction of searching for a data stream dividing point, wherein N*U is not greater than $\| B_z \| + \max_x(\| A_x \| + \| (k_i - p_{ix}) \|)$, so as to obtain a new potential dividing point, and performing step (a); and

(c) when at least a part of data in each window $W_{ix}[p_{ix} - A_x, p_{ix} + B_x]$ of M windows of the current potential dividing point k_i meets the preset condition C_x, selecting the current potential dividing point k_i as a data stream dividing point.

18. The server according to claim 17, wherein the rule further comprises that at least two points p_e and p_f meet conditions $A_e = A_f$, $B_e = B_f$, and $C_e = C_f$.

19. The server according to claim 18, wherein the rule further comprises: relative to the potential dividing point k, the at least two points p_e and p_f are in a direction opposite to the direction of searching for a data stream dividing point.

20. The server according to claim 18 or 19, wherein the rule further comprises that a distance

between the at least two points p_e and p_f is 1 U.

21. The server according to any one of claims 17 to 20, wherein the central processing unit is specifically configured to determine, by using a random function, whether the at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ meets the preset condition C_z.

22. The server according to claim 21, wherein the central processing unit is specifically configured to determine, by using a hash function, whether the at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ meets the preset condition C_z.

23. The server according to any one of claims 17 to 22, wherein when the at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the point p_{iz} along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[\ p_{ic}-A_c,\ p_{ic}+B_c\ ]$ corresponding to a point p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ or a left boundary of a window $W_{ic}[\ p_{ic}-A_c,\ p_{ic}+B_c\ ]$ corresponding to a point p_{ic} that is determined for the new potential dividing point falls within a range of the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$, wherein the point p_{ic} determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to the ~~a data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

24. The server according to claim 21, wherein the determining, by the central processing unit by using a random function, whether the at least a part of data in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$ meets the preset condition C_z specifically comprises:

selecting F bytes in the window $W_{iz}[\ p_{iz}-A_z,\ p_{iz}+B_z\ ]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, wherein F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, wherein when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, wherein $a_{m,n}$ represents any one of $a_{m,1}$, ..., and

$a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + \ldots + V_{am,8} * h_{m,8}$, S_{a1}, S_{a2}, ..., and S_{aF*H} are obtained in a same way, a quantity K of values greater than 0 among S_{a1}, S_{a2}, ..., and S_{aF*H} is counted, and when K is an even number, the at least a part of data in the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ meets the preset condition C_z.

25. A server for searching for a data stream dividing point, wherein the server comprises a central processing unit and a primary memory, the central processing unit communicates with the primary memory, and a rule preset on the server~~a rule is preset on the server, wherein the rule~~ is: for a potential dividing point k, determining M windows $W_x[k - A_x, k + B_x]$ and a preset condition C_x corresponding to the window $W_x[k - A_x, k + B_x]$, wherein x indicates consecutive natural numbers from 1 to M, M≥2, and A_x and B_x are integers; and

the main memory is configured to store an executable instruction, and the central processing unit executes the executable instruction to perform the following steps:

(a) determining a corresponding window $W_{iz}[k_i - A_z, k_i + B_z]$ for a current potential dividing point k_i according to the rule, wherein i and z are integers, and 1≤z≤M;

(b) determining whether at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets a preset condition C_z; and

when the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ does not meet the preset condition C_z, skipping N minimum units U for searching for a data stream dividing point from the current potential dividing point k_i along a direction of searching for a data stream dividing point,

wherein N*U is not greater than $\| B_z \| + \max_x(\| A_x \|)$, so as to obtain a new potential dividing point, and performing step (a); and

(c) when at least a part of data in each window $W_{ix}[k_i - A_x, k_i + B_x]$ of M windows of the current potential dividing point k_i meets the preset condition C_x, selecting the current potential dividing point k_i as a data stream dividing point.

26. The server according to claim 25, wherein the rule further comprises that at least two windows $W_{ie}[k_i - A_e, k_i + B_e]$ and $W_{if}[k_i - A_f, k_i + B_f]$ meet conditions $|A_e + B_e| = |A_f + B_f|$ and $C_e = C_f$.

27. The server according to claim 26, wherein ~~a rule is preset for the server, and~~ the rule further comprises that A_e and A_f are positive integers.

28. The server according to claim 26 or 27, wherein the rule further comprises $A_e - 1 = A_f$ and $B_e + 1 = B_f$.

29. The server according to any one of claims 25 to 28, wherein the central processing unit is specifically configured to determine, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z.

30. The server according to claim 29, wherein the central processing unit is specifically configured to determine, by using a hash function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z.

31. The server according to any one of claims 25 to 30, wherein when the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the current potential dividing point k_i along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[k_i - A_z, k_i + B_z]$ or a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point falls within a range of the window $W_{iz}[k_i - A_z, k_i + B_z]$, wherein the window $W_{ic}[k_i - A_c, k_i + B_c]$ determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to ~~a~~ the ~~data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the

rule.

32. The server according to claim 29, wherein the determining, by the central processing unit by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z specifically comprises:

selecting F bytes in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, wherein F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, wherein when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, wherein $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + \ldots + V_{am,8} * h_{m,8}$, S_{a1}, S_{a2}, ..., and S_{aF*H} are obtained in a same way, a quantity K of values greater than 0 among S_{a1}, S_{a2}, ..., and S_{aF*H} is counted, and when K is an even number, the at least a part of data in the window $W_{iz}[k_i - A_z,\ k_i + B_z]$ meets the preset condition C_z.

33. A server for searching for a data stream dividing point, wherein a rule preset on the server ~~a rule is preset on the server, wherein the rule~~ is: for a potential dividing point k, determining M points p_x, a window $W_x[p_x - A_x,\ p_x + B_x]$ corresponding to the point p_x, and a preset condition C_x corresponding

unit is configured to: when the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ does not meet the preset condition C_z, skip the N minimum units U for searching for a data stream dividing point from the point p_{iz} along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, the determining unit performs step (a) for the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ or a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ ~~that is determined for the new potential dividing point~~ falls within a range of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$, wherein the point p_{ic} ~~window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$~~ determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to ~~a~~ the ~~data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

40. The server according to claim 37, wherein that the judging and processing unit is specifically configured to determine, by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z specifically comprises:

selecting F bytes in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, wherein F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, wherein when $a_{m,n}=1$, $V_{am,n}=1$, and when $a_{m,n}=0$, $V_{am,n}=-1$, wherein $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as:

$\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + \ldots + V_{am,8} * h_{m,8}$, S_{a1}, S_{a2}, ..., and S_{aF*H} are obtained in a same way, a quantity K of values greater than 0 among S_{a1}, S_{a2}, ..., and S_{aF*H} is counted, and when K is an even number, the at least a part of data in the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ meets the preset condition C_z.

41. A server for searching for a data stream dividing point, wherein a rule ~~is~~ preset on the server~~, wherein the rule~~ is: for a potential dividing point k, determining M windows $W_x[k - A_x, k + B_x]$ and a preset condition C_x corresponding to the window $W_x[k - A_x, k + B_x]$, wherein x indicates consecutive natural numbers from 1 to M, M≥2, and A_x and B_x are integers; and

the server comprises: a determining unit, configured to perform step (a):

(a) determining a corresponding window $W_{iz}[k_i - A_z, k_i + B_z]$ for a current potential dividing point k_i according to the rule, wherein i and z are integers, and 1≤z≤M; and

a judging and processing unit, configured to determine whether at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets a preset condition C_z;

when the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ does not meet the preset condition C_z, skip N minimum units U for searching for a data stream dividing point from the current potential dividing point k_i along a direction of searching for a data stream dividing point, wherein N*U is not greater than $\| B_z \| + \max_x(\| A_x \|)$, so as to obtain a new potential dividing point, wherein the determining unit performs step (a) for the new potential dividing point; and

when at least a part of data in each window $W_{ix}[k_i - A_x, k_i + B_x]$ of M windows of the current potential dividing point k_i meets the preset condition C_x, select the current potential dividing point k_i as a data stream dividing point.

42. The server according to claim 41, wherein the rule further comprises that at least two windows $W_{ie}[k_i - A_e, k_i + B_e]$ and $W_{if}[k_i - A_f, k_i + B_f]$ meet conditions $|A_e + B_e| = |A_f + B_f|$ and $C_e = C_f$.

43. The server according to claim 42, wherein the rule further comprises that A_e and A_f are positive integers.

44. The server according to claim 42 or 43, wherein the rule further comprises A_e-1= A_f and B_e+1= B_f.

45. The server according to any one of claims 41 to 44, wherein the judging and processing unit is specifically configured to:

determine, by using a random function, whether the at least a part of data in the window W_{iz} [k_i - A_z, k_i + B_z] meets the preset condition C_z.

46. The server according to claim 45, wherein the judging and processing unit specifically determines, by using a hash function, whether the at least a part of data in the window W_{iz} [k_i - A_z, k_i + B_z] meets the preset condition C_z.

47. The server according to any one of claims 41 to 46, wherein the judging and processing unit is configured to: when the at least a part of data in the window W_{iz} [k_i - A_z, k_i + B_z] does not meet the preset condition C_z, skip the N minimum units U for searching for a data stream dividing point from the current potential dividing point k_i along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, the determining unit performs step (a) for the new potential dividing point, and according to the rule, a left boundary of a window W_{ic} [k_i - A_c, k_i + B_c] that is determined for the new potential dividing point coincides with a right boundary of the window W_{iz} [k_i - A_z, k_i + B_z] or a left boundary of a window W_{ic} [k_i - A_c, k_i + B_c] that is determined for the new potential dividing point falls within a range of the window W_{iz} [k_i - A_z, k_i + B_z], wherein the window W_{ic} [k_i - A_c, k_i + B_c] determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to ~~a~~ the ~~data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the rule.

48. The server according to claim 46, wherein that the judging and processing unit determines, by using a random function, whether the at least a part of data in the window W_{iz} [k_i - A_z, k_i + B_z] meets the preset condition C_z specifically comprises:

selecting F bytes in the window W_{iz} [k_i - A_z, k_i + B_z], and using the F bytes repeatedly H times to obtain F*H bytes in total, wherein F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits

corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, wherein when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, wherein $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + \ldots + V_{am,8} * h_{m,8}$, S_{a1}, S_{a2}, ..., and S_{aF*H} are obtained in a same way, a quantity K of values greater than 0 among S_{a1}, S_{a2}, ..., and S_{aF*H} is counted, and when K is an even number, the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z.

49. A computer-readable storage medium, wherein the computer-readable storage medium is configured to store an executable instruction, a server executes the executable instruction to search for a data stream dividing point, and a rule ~~is~~ preset on the server~~, wherein the rule~~ is: for a potential dividing point k, determining M points p_x, a window $W_x[p_x - A_x, p_x + B_x]$ corresponding to the point p_x, and a preset condition C_x corresponding to the window $W_x[p_x - A_x, p_x + B_x]$, wherein x indicates consecutive natural numbers from 1 to M, M≥2, and A_x and B_x are integers; and

the server executes the executable instruction to perform the following steps:

(a) determining p_{iz} and a window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ corresponding to p_{iz} for a current potential dividing point k_i according to the rule, wherein i and z are integers, and 1≤z≤M;

(b) determining whether at least a part of data in the window $W_{iz}[p_{iz} - A_z, p_{iz} + B_z]$ meets a

from the point p_{iz} along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ or a left boundary of a window $W_{ic}[p_{ic}-A_c,\ p_{ic}+B_c]$ corresponding to a point p_{ic} that is determined for the new potential dividing point falls within a range of the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$, wherein the point p_{ic} determined for the new potential dividing point is a point ranking the first in a sequence, which is obtained according to ~~a the data stream search direction~~direction of searching for a data stream dividing point, of M points that are determined for the new potential dividing point according to the rule.

56. The computer-readable storage medium according to claim 53, wherein the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$ meets the preset condition C_z specifically comprises:

selecting F bytes in the window $W_{iz}[p_{iz}-A_z,\ p_{iz}+B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, wherein F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits corresponding to the F*H bytes can be represented as: $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,8} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ a_{F*H,1} & a_{F*H,2} & \cdots & a_{F*H,8} \end{pmatrix}$, wherein when $a_{m,n}$=1, $V_{am,n}$=1, and when $a_{m,n}$=0, $V_{am,n}$=-1, wherein $a_{m,n}$ represents any one of $a_{m,1}$, ..., and $a_{m,8}$, a matrix V_a is obtained according to a conversion relationship between $a_{m,n}$ and $V_{am,n}$ from the bits corresponding to the F*H bytes, the matrix V_a is represented as: $\begin{pmatrix} V_{a1,1} & V_{a1,2} & \cdots & V_{a1,8} \\ V_{a2,1} & V_{a2,2} & \cdots & V_{a2,8} \\ \vdots & \vdots & \vdots & \vdots \\ V_{aF*H,1} & V_{aF*H,2} & \cdots & V_{aF*H,8} \end{pmatrix}$, F*H*8 random numbers are selected from random numbers following normal distribution to form a matrix R, the matrix R is represented as: $\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,8} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,8} \\ \vdots & \vdots & \vdots & \vdots \\ h_{F*H,1} & h_{F*H,2} & \cdots & h_{F*H,8} \end{pmatrix}$, random numbers in an mth row of the matrix V_a and an mth row of

the matrix R are multiplied and products are added to obtain a value, which is specifically represented as $S_{am} = V_{am,1} * h_{m,1} + V_{am,2} * h_{m,2} + ... + V_{am,8} * h_{m,8}$, S_{a1}, S_{a2}, ..., and S_{aF*H} are obtained in a same way, a quantity K of values greater than 0 among S_{a1}, S_{a2}, ..., and S_{aF*H} is counted, and when K is an even number, the at least a part of data in the window $W_{iz}[\, p_{iz} - A_z,\ p_{iz} + B_z\,]$ meets the preset condition C_z.

57. A computer-readable storage medium, wherein the computer-readable storage medium is configured to store an executable instruction, a server executes the executable instruction to search for a data stream dividing point, and a rule ~~is~~ preset on the server~~, wherein the rule~~ is: for a potential dividing point k, determining M windows $W_x[\, k - A_x,\ k + B_x\,]$ and a preset condition C_x corresponding to the window $W_x[\, k - A_x,\ k + B_x\,]$, wherein x indicates consecutive natural numbers from 1 to M, M≥2, and A_x and B_x are integers; and

the server executes the executable instruction to perform the following steps:

(a) determining a corresponding window $W_{iz}[\, k_i - A_z,\ k_i + B_z\,]$ for a current potential dividing point k_i according to the rule, wherein i and z are integers, and 1≤z≤M;

(b) determining whether at least a part of data in the window $W_{iz}[\, k_i - A_z,\ k_i + B_z\,]$ meets a preset condition C_z; and

when the at least a part of data in the window $W_{iz}[\, k_i - A_z,\ k_i + B_z\,]$ does not meet the preset condition C_z, skipping N minimum units U for searching for a data stream dividing point from the current potential dividing point k_i along a direction of searching for a data stream dividing point, wherein N*U is not greater than $\| B_z \| + \max_x(\| A_x \|)$, so as to obtain a new potential dividing point, and performing step (a); and

(c) when at least a part of data in each window $W_{ix}[\, k_i - A_x,\ k_i + B_x\,]$ of M windows of the current potential dividing point k_i meets the preset condition C_x, selecting the current potential dividing point k_i as a data stream dividing point.

58. The computer-readable storage medium according to claim 57, wherein the rule further comprises that at least two windows $W_{ie}[\, k_i - A_e,\ k_i + B_e\,]$ and $W_{if}[\, k_i - A_f,\ k_i + B_f\,]$ meet conditions $|\, A_e + B_e \,| = |\, A_f + B_f \,|$ and $C_e = C_f$.

59. The computer-readable storage medium according to claim 58, wherein the rule further comprises that A_e and A_f are positive integers.

60. The computer-readable storage medium according to claim 58 or 59, wherein the rule further comprises A_e-1= A_f and B_e+1= B_f.

61. The computer-readable storage medium according to any one of claims 57 to 60, wherein:

the determining, by the server, whether at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets a preset condition C_z specifically comprises:

determining, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z.

62. The computer-readable storage medium according to claim 61, wherein the determining, by the server by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z is specifically: determining, by the server by using a hash function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z.

63. The computer-readable storage medium according to any one of claims 57 to 62, wherein when the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ does not meet the preset condition C_z, the N minimum units U for searching for a data stream dividing point are skipped from the current potential dividing point k_i along the direction of searching for a data stream dividing point, so as to obtain the new potential dividing point, and according to the rule, a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point coincides with a right boundary of the window $W_{iz}[k_i - A_z, k_i + B_z]$ or a left boundary of a window $W_{ic}[k_i - A_c, k_i + B_c]$ that is determined for the new potential dividing point falls within a range of the window $W_{iz}[k_i - A_z, k_i + B_z]$, wherein the window $W_{ic}[k_i - A_c, k_i + B_c]$ determined for the new potential dividing point is a window ranking the first in a sequence, which is obtained according to ~~a the data stream search direction~~direction of searching for a data stream dividing point, of M windows that are determined for the new potential dividing point according to the rule.

64. The computer-readable storage medium according to claim 61, wherein the determining, by using a random function, whether the at least a part of data in the window $W_{iz}[k_i - A_z, k_i + B_z]$ meets the preset condition C_z specifically comprises:

selecting F bytes in the window $W_{iz}[k_i - A_z, k_i + B_z]$, and using the F bytes repeatedly H times to obtain F*H bytes in total, wherein F≥1, each byte is formed by 8 bits, which are denoted as $a_{m,1}$, ..., and $a_{m,8}$, representing the 1st bit to the 8th bit of an mth byte in the F*H bytes, bits