CLAIMS

1. A method for identifying an applet of risky content based on differential privacy preserving, applied to a client of a service platform and comprising:

in response to monitoring a target operation performed by a user by using an applet facilitated by the service platform, obtaining a result of determining whether an operation source is a risky operation source based, at least in part, on a risk identification policy group, wherein the operation source includes a combination of the user and a device used by the user;

in response to determining that the operation source is a risky operation source, disturbing a program identifier of the applet by using a randomized response technique that satisfies one or more criteria of local differential privacy preserving, to obtain a disturbed program identifier; and

sending an operation source identifier of the operation source and the disturbed program identifier to a serving end of the service platform, so that the serving end identifies whether the applet is an applet of risky content based on a plurality of received disturbed program identifiers by using the randomized response technique, wherein the operation source identifier includes an account identifier and a device identifier, and

wherein the risk identification policy group includes a content risk identification policy being used for risk identification for at least one of (a) information content related to the target operation in the applet or (b) historical data corresponding to the operation source identifier.

2. The method according to claim 1, further comprising:

before responding to the monitoring of the target operation performed by the user by using the applet carried by the service platform,

receiving the risk identification policy group from the serving end.

3. The method according to claim 1, wherein obtaining the result of determining whether the operation source is a risky operation source includes:

determining whether the operation source has a risk mark, the risk mark being a mark added in advance for a risky operation source based, at least in part, on the risk identification policy group; and

in response to determining that the operation source has the risk mark, determining that the operation source is a risky operation source.

4. The method according to claim 1, wherein obtaining the result of determining whether the operation source is a risky operation source includes:

determining whether the operation source is a risky operation source by using the risk identification policy group.

5. The method according to claim 4, further comprising:

after responding to determining that the operation source is a risky operation source, adding a risk mark to the operation source.

6. The method according to claim 1, wherein the target operation includes at least one of an information release operation or an information access operation.

7. The method according to claim 1,

the risk identification policy group includes at least one of a content risk identification policy, a device risk identification policy, or a list policy, the list policy including at least one of an account list or a device list;

the device risk identification policy is used for risk identification for the device corresponding to the device identifier; and

the list policy is used for risk identification for the device corresponding to at least one of the device identifier or an account corresponding to the account identifier.

8. The method according to claim 4, wherein determining whether the operation source is a risky operation source by using the risk identification policy group includes:

determining a target risk value corresponding to the operation source by using the risk identification policy group; and performing:

in response to determining that the target risk value exceeds a risk threshold, determining that the operation source is a risky operation source; or

in response to determining that the target risk value does not exceed the risk threshold, determining that the operation source is not a risky operation source.

9. The method according to claim 8, wherein determining the target risk value corresponding to the operation source by using the risk identification policy group includes:

in response to determining that the risk identification policy group includes a single risk identification policy, determining the target risk value by using the single risk identification policy; or

in response to determining that the risk identification policy group includes a plurality of risk identification policies, determining a plurality of risk values for the operation source by using the plurality of risk identification policies, and determining the target risk value based on the plurality of risk values.

10. The method according to claim 9, wherein determining the target risk value based on the plurality of risk values includes:

determining at least one of an average value or a sum of the plurality of risk values as the target risk value.

11. The method according to claim 9, wherein individual policies in the risk identification policy group are associated with respective weights, and determining the target risk value based on the plurality of risk values includes:

performing weighted summation on the plurality of risk values based, at least in part, on the weights associated with the policies in the risk identification policy group, and determining the target risk value based, at least in part, on the weighted summation.

12. The method according to claim 8, wherein determining the target risk value corresponding to the operation source by using the risk identification policy group includes:

in response to determining that the risk identification policy group includes a content risk identification policy, obtaining at least one of (a) information content related to the target operation in the applet or (b) historical data corresponding to the operation source identifier; and

determining the target risk value based, at least in part, on the target data by using the content risk identification policy.

13. The method according to claim 12, wherein the content risk identification policy includes a keyword policy and a content risk identification model, the keyword policy includes at least one keyword and a risk value corresponding to the at least one keyword, and the content risk identification model is used to identify a risk value corresponding to input content, and wherein determining the target risk value based, at least in part, on the target data by using the risk identification policy group includes:

in response to determining that the target data includes a target keyword that matches the at least one keyword, determining the target risk value based, at least in part, on a risk value corresponding to each target keyword; or

in response to determining that the target data does not include the target keyword, identifying a risk value corresponding to the target data by using the content risk identification model, and determining the target risk value based, at least in part, on the risk value.

14. The method according to claim 8, wherein the operation source identifier includes an account identifier and a device identifier, and wherein determining the target risk value corresponding to the operation source includes:

in response to determining that the risk identification policy group including a list policy, determining whether at least one information item in the operation source identifier is included in the list policy, wherein the list policy includes at least one of a device list or an account list, the device list and the account list are associated with respective risk values, the device list includes at least one device identifier, and the account list includes at least one account identifier; and

in response to determining that the at least one information item is included in the list policy, determining the target risk value based, at least in part, on a risk value corresponding to the at least one information item.

15. The method according to claim 1, wherein disturbing the program identifier of the applet includes:

converting the program identifier of the applet into a binary string of a target length; and

disturbing the binary string based, at least in part, on a disturbing probability by using the randomized response technique.

16. The method according to claim 1, wherein the randomized response technique includes any one k-RR algorithm, RAPPOR algorithm, or CMS algorithm.