一种基于规则的安全日志范式处理模型

王倩　陆晟　龚俭
(东南大学计算机科学与工程系　　南京　　210096)

【摘要】通过日志处理，能够有效掌握系统运行情况、加强系统的维护与管理。由于系统日志数据量大，种类众多，语法和语义各不相同，一般不直接对原始日志进行处理。本文以安全日志分析为例，介绍了日志范式处理的思想，提出了一种基于规则的安全日志范式处理模型，并给出了实现示例以及进一步的工作方向。

【关键词】范式，系统安全，日志

一、引言

系统日志记录了系统运行的各种原始信息。通过对这些信息的统计、分析、与综合，能够有效地掌握系统运行情况、诊断差错事故、发现安全事件、了解流量分布等等，从而加强系统的维护与管理。因此，在网管、安全、计费等领域中，都需要进行日志处理，并对系统日志加以保存。

由于系统日志种类众多，数据量大，假如直接对原始日志进行处理，工作量大，程序运行效率低，保存日志记录需要占用大量存储空间。而且，原始日志的语法和语义各不相同，只能针对每种日志开发专门的处理程序，这将会带来大量的冗余工作量，程序的灵活性与适应性都较差，而且难以实现不同日志之间相关信息的综合处理。因此，一般不会直接对系统提供的原始日志进行处理。本文以安全日志为例，介绍了一种系统日志的范式处理方法。

二、系统日志的范式处理

事实上，特定的应用领域所需要的日志信息是一定的，并不需要系统日志提供的全部内容，而只关心其中的某些元素。例如，网管领域侧重于日志记录中的统计型信息，而安全领域侧重事件型信息。因此，当面向特定目的的应用时，可以定义一个相对固定的范式形式，在一定的政策控制下对原始日志信息进行加工，抽取所需要的元素，滤掉无关成份，映射为统一的范式表达，以一致的语法和语义形式屏蔽原始日志的不同形式，在此基础上进行统一地处理。这种方法即称为系统日志的范式处理。

采用范式处理，减少了逐一开发日志处理程序的工作量，并能综合来自不同日志的信息，具有较广的适应性和较强的灵活性。同时大大减少了处理数据量，提高处理效率，仅保存范式记录也节省了系统存储空间。

三、安全日志的范式处理

本文中所称的安全日志指安全领域所涉及的系统日志，如各种 login 日志、ftp 日志、smtp 日志等。对安全日志的分析与综合处理，是发现安全事件、洞察系统漏洞、评判系统安全程度、进而加强系统安全管理的重要手段。

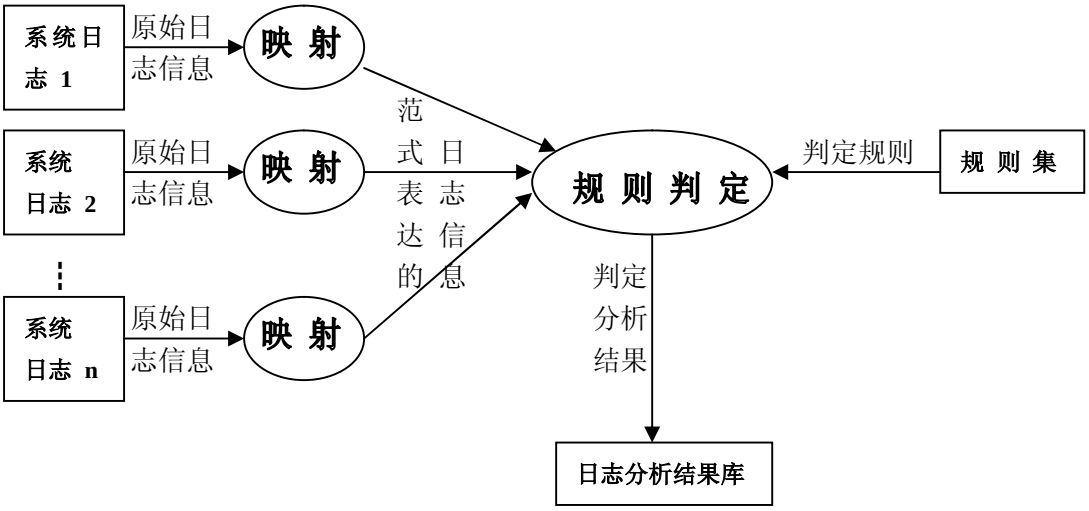
安全问题一般表现为事件形式，所以对安全日志的分析侧重于日志记录中的事件信息，即何时、何地、何主体对何客体进行了何种操作，而不关心其中的统计信息。例如，对于 ftp 日志，需要何时、何人、自何源点、从/向何终点、下载/上载何文件等事件型信息，而不关心传输文件数、传输文件长度等统计型信息。因此，安全日志的范式定义是在去除各种统计型信息之后，对所需的事件型信息的语法、语义定义。

将安全日志映射为一致的范式表达，即可在此基础上进行自动地统一分析和综合处理。本文以下即提出了一种基于规则的安全日志范式处理模型。

四、基于规则的安全日志范式处理模型

1、基本思想

安全日志分析的目的在于判定系统事件的安全性、发现安全事件。对安全性的评判基于一组定义的规则，这组规则针对日志记载的各种系统行为与事件分别给出其判定条件与相应结论。因为所有日志映射为同一范式表达，以此为基础，针对不同系统日志的所有规则都能够以一致的形式描述，从而能对来自不同原始日志的条目同一地加以判定。将映射为范式表达的安全日志通过所有规则统一地、综合地判定，产生分析结果。这一模型可用如下的数据流图表示。



2、日志的范式表达

因为安全分析多侧重于事件形式，所以安全日志的范式定义一般应具有以下事件信息：时间、地点、主体、活动。这些信息还可以进一步细化：时间可细分为到达时间和离开时间，地点可以进一步划分为源地址和宿地址，活动可分解为客体和访问动作，等等。因此，日志范式的格式可简单表示为：

(主体名,(到达时间,离开时间),(源地址,宿地址),(客体名,访问动作))。

对于安全性分析，这些信息已较为完备了。但有时候，判定规则可能要求更多信息元素。比如，某些系统将传送过大的文件视为安全事件，这时即需要传输文件长度信息。因此日志的范式表达应定义为相对固定的、可扩充的形式，例如采用 ASN.1 的描述方式。

3、规则的范式表达

判定规则表述为判定条件与相应的逻辑结论，而判定条件一般是针对事件元素（时间、地点等等）及其综合关系定义。由于所有日志映射为一致的范式表达，所以对元素的条件判定可用统一定义的谓词描述和处理。又因为日志范式元素相对固定，这些谓词组成一个相对确定的谓词集，即可在此谓词集基础上定义规则范式。

采用谓词逻辑的方法，用自然语言描述的规则可表示为一形如

条件谓词公式 $\Rightarrow$ 结论谓词

的逻辑表达式，其条件以谓词集中相应谓词组成的逻辑公式加以描述。而条件谓词公式总能

化为标准型的前束合取范式，所以可将其进一步简化表示为子句集的形式，即：

（条件谓词子句集）⇒ 结论谓词，

条件子句以统一定义的谓词集为基，故而具有统一的语法、语义形式，作为判定规则的范式表达。

4、规则判定过程

（1）三值逻辑真值计算规则补充

并非全部系统日志都具有范式中定义的所有元素，因此，日志条目可能会有部分空缺，使判定规则的某些条件子句无法确定。因而，逻辑判定过程应使用三值逻辑：T（true）、F（false）、U（undefined），需要补充定义如下真值计算规则：

① U∧T＝T∧U＝U　② U∧F＝F∧U＝F　③ U∧U＝U

④ U∨T＝T∨U＝T　⑤ U∨F＝F∨U＝U　⑥ U∨U＝U　⑦┐U＝U

这样即可进行三值逻辑基础上的日志分析。

（2） 判定规则的结论获得

基于规则的范式定义，对单条规则结论的推导过程可转化为条件子句集可满足性的逻辑判定问题，采用人工智能的有关技术自动实现。当日志条目满足条件子句集时，即推出结论，记录结果信息，触发有关处理动作。

可对规则设定优先级，当同一条或几条日志条目推出多项结论时，依据优先级政策决定最终结论。

五、实现方案示例

本节以安全日志分析的一个子集系统为例，说明上文所论模型的实现。

1、系统描述

本例对 Unix 系统的 wtmp 与 sulog 两类日志进行基于以下规则的安全分析：

规则 1、如果除 billy 之外的其它户头成功 su 为 root，则超级用户口令失密；

规则 2、如果有帐户从 IP 地址 202.112.23.0 - 202.112.23.255 以外的主机成功登录，则该帐户被攻破；

规则 3、如果有帐户从 IP 地址 202.112.23.0 - 202.112.23.255 以外主机成功登录且成功 su 为 root，则系统被攻破。

规则优先级依次从低到高。推出多条结论时，以最高优先级的规则结论作为最终结论。

2、日志范式表达

本系统中日志范式的格式可简单表示为：

(主体名,(到达时间,离开时间),(源地址,宿地址),(客体名,访问动作))。

设主机 hanan 上有如下两条系统日志：

① wtmp 日志

billy pts/22 202.112.25.213 Thu Jul 2 11:44 - 12:16 (00:31)

② sulog 日志

SU 07/02 12:01 + pts/22 billy-root

将其映射为以下的范式表达：

Log1：

(billy,(07.02.11:44,07.02.12:16),(202.112.25.213,hanan:22),(,login))

Log2：

(billy,（07.02.12:01,),（, hanan:22),（root, su))

这样，即将不同语法和语义的日志条目规范为统一的范式形式。

3、规则范式表达

根据日志范式形式，可定义以下谓词集：

U(x,L) -- x 是日志条目 L 的"主体名"元素

AT(x,L) -- x 是日志条目 L 的"到达时间"元素

LT(x,L) -- x 是日志条目 L 的"离开时间"元素

S(x,L) -- x 是日志条目 L 的"源地址"元素

D(x,L) -- x 是日志条目 L 的"宿地址"元素

O(x,L) -- x 是日志条目 L 的"客体名"元素

A(x,L) -- x 是日志条目 L 的"访问动作"元素

Is(x,y) -- x 是 y

Between(x,x1,x2) -- x 在 x1 与 x2 之间

定义结论谓词：

P(L) --日志条目 L 为超户口令失密事件；

Q(L) --日志条目 L 表示帐户被攻破；

R(L1,L2) --日志条目 L1、L2 表示系统被攻破。

则本系统的三条规则可表示为如下逻辑表达式：

Rule1'：

$\exists L \forall x((U(x,L) \supset \neg Is(x,billy)) \wedge (O(x,L) \supset Is(x,root)) \wedge (A(x,L) \supset Is(x,su))) \Rightarrow P(L)$

Rule2'：

$\exists L \forall x((S(x,L) \supset \neg Between(x,202.112.23.0,202.112.23.255)) \wedge (A(x,L) \supset Is(x,login))) \Rightarrow Q(L)$

Rule3'：

$\exists L1 \exists L2 \forall x \forall y \forall z(((U(x,L1) \wedge U(y,L2)) \supset Is(x,y)) \wedge ((S(x,L1) \wedge S(y,L2)) \supset Is(x,y)) \wedge ((D(x,L1) \wedge D(y,L2)) \supset Is(x,y)) \wedge ((AT(x,L1) \wedge LT(z,L1) \wedge AT(y,L2)) \supset Between(y,x,z)) \wedge (S(x,L1) \supset \neg Between(x,202.112.23.0,202.112.23.255)) \wedge (O(y,L2) \supset Is(y,root)) \wedge (A(x,L1) \supset Is(x,login)) \wedge (A(y,L2) \supset Is(y,su))) \Rightarrow R(L1,L2)$

将条件表示为子句集形式，则化为如下规则范式：

Rule1：

$(\neg U(x,L) \vee \neg Is(x,billy), \neg O(x,L) \vee Is(x,root), \neg A(x,L) \vee Is(x,su)) \Rightarrow P(L)$

Rule2：

$(\neg S(x,L) \vee \neg Between(x,202.112.23.0,202.112.23.255), \neg A(x,L) \vee Is(x,login)) \Rightarrow Q(L)$

Rule3：

$(\neg U(x,L1) \vee \neg U(y,L2) \vee Is(x,y), \neg S(x,L1) \vee \neg S(y,L2) \vee Is(x,y), \neg D(x,L1) \vee \neg D(y,L2) \vee Is(x,y), \neg AT(x,L1) \vee \neg LT(z,L1) \vee \neg AT(y,L2) \vee Between(y,x,z), \neg S(x,L1) \vee \neg Between(x,202.112.23.0,202.112.23.255), \neg O(y,L2) \vee Is(y,root), \neg A(x,L1) \vee Is(x,login), \neg A(y,L2) \vee Is(y,su)) \Rightarrow R(L1,L2)$

4、规则判定

采用人工智能中谓词逻辑判定的方法，Log1、Log2 经 Rule1 - Rule3 的自动判定，推出结论：Q(Log1)、R(Log1,Log2)。按照优先级政策，最终结论为 R(Log1,Log2)，即系统被攻破。这时将 Log1、Log2 内容作为结果参数记录下来，并激发相应处理。

六、结论

本文以安全日志分析为例，介绍了日志范式处理的思想，提出了一种基于规则的安全日志范式处理模型，并给出了实现示例。该系统的处理能力在很大程度上取决于规则集的完备性。目前，这一系统尚有许多待完善的工作。比如，形式刻画日志映射的政策控制模型；通过定义逻辑推演规则，实现判定规则的自动生成；在元数学层次上对规则集无矛盾性与完备性检查等等，以进一步提高日志处理的形式化和智能化水平。

七、参考文献

[1] 《Solaris 2.4：Security，Performance and Accounting Administration》
[2] 莫绍揆 著《数理逻辑教程》 华中工学院出版社
[3] 杨祥金 蔡庆生 编著《人工智能》 科学技术文献出版社重庆分社

A Rules Based Normal-formed Processing Model of Security Logs

Wang Qian　Lu Sheng　Gong Jian
(Computer Science and Engineering Department,Southeast University　Nanjing　210096)

【Abstract】With the processing of system logs, we can efficiently master the system running states and strengthen the system maintenance and management. As the information contained in the original system logs is too much and too complex to be processed directly, a normal-formed processing method for security logs is given out in this paper, which uses a set of normal-formed rules for modeling the processing.　A example is arisen also in detail for explaining how the model works.

【Keywords】normal-form,　system security,　logs