第32卷 第17期 计 算 机 工 程 2006年9月
Vol.32 № 17 Computer Engineering September 2006
·发展趋势/热点技术· 文章编号：1000—3428(2006)17—0001—03 文献标识码：A 中图分类号：TP393.08

网络安全信息关联与分析技术的研究进展

彭雪娜，闻英友，赵 宏
（东北大学计算机软件国家工程研究中心，沈阳 110004）

摘 要：介绍了网络安全信息关联分析技术的背景，指出了该技术解决的问题。根据分析方法的不同，将该技术的现有方法分为4类：基于网络安全信息相似性的分析技术，基于攻击场景识别的分析技术，基于网络安全信息因果关系的分析技术，基于网络安全信息统计因果关系的分析技术。对每类方法的基本思想、现有技术以及存在的问题进行了阐述和分析，对未来的一些工作方向进行了展望。
关键词：网络安全；信息分析；告警聚集；告警关联

A Survey on Network Security Information Correlation Techniques

PENG Xuena, WEN Yingyou, ZHAO Hong
(State Engineering Research Center of Computer Software, Northeastern University, Shenyang 110004)

【Abstract】 The background of the network security information correlation technique is introduced, and the problem that it supposed to solve is clarified. And according to the different methods used in the technique, this paper classifies the methods into four categories: similarity-based analysis, attack-scenario based analysis, causality knowledge based analysis and statical causality based anaylsis. For each category, the basic idea and the existing techniques are introduced and analyzed, and the unsolved problems are pointed out. The development direction and future works are analyzed.
【Key words】 Network security; Information analysis; Alert clustering; Alert correlation

随着网络技术的发展和应用的迅速普及，网络安全越来越受到人们的重视。然而，大规模网络安全事件的产生以及众多网络安全设备的报警信息，对网络安全管理提出了严峻的挑战。如何实现对网络安全信息的高效分析和提取，保障信息网络的整体安全已成为当前国内外网络安全技术领域亟待解决的重要问题。网络安全信息分析技术是解决该问题的一项重要技术，主要研究如何从海量网络安全信息中提取出真正有价值的安全信息。该技术涉及的具体内容包括对来自多种部件的网络安全信息的统一管理和分析，对当前网络中存在或隐藏的安全问题的发掘，对网络整体安全态势的合理评估和预警。由于对网络安全信息分析技术的研究具有较高的理论研究价值和实际应用价值，近年来，该领域已经成为一个研究热点。

1 网络安全信息分析的难点及技术概述

传统网络安全信息分析和维护通常是由网络管理员人工完成。随着各种网络安全事件的大规模产生，由管理员手工进行的网络安全信息处理和维护已经不能满足网络安全管理的实际需要，主要表现为：(1)传统网络安全信息分析由各安全部件分散维护和管理，信息格式和语义不统一，部件间缺乏信息共享。(2)传统网络安全信息分析主要依赖手工操作，由于部件产生的安全信息本身误报率和漏报率较高、数据量大、信息间缺乏关联，因此安全信息分析工作效率低、成本高、效果差。问题(1)主要涉及网络安全信息的采集和规范化问题，由于不作为本文关注点，因此不做详细阐述，建议查看IETF的IDWG提出的规范建议资料。问题(2)主要涉及网络安全信息的分析问题。解决该问题涉及的相关技术较多，网络安全信息关联分析技术是解决该问题的一项重要技术，该技术主要解决网络安全信息数据量大、语义级别低、信息间缺乏关联等问题，旨在发现安全信息之间的联系并对特定攻击场景进行识别和抽象。本文将主要阐述此项技术的研究现状，并对网络安全信息分析技术的整体发展趋势进行展望。

2 网络安全信息关联分析技术研究进展

2.1 基于网络安全信息相似性的分析技术

基于网络安全信息相似性的分析技术源于同源触发的网络安全信息具有相似性的观察结果，并基于相似网络安全信息源于相同或相似根源的假设，来实现对网络安全信息关系和攻击场景的识别。

基于网络安全信息相似性的分析技术所采用的最主要的方法是基于概率聚类技术的分析方法[1]，由Alfonso Valdes首次提出。基于概率聚类的网络安全信息分析方法是通过计算网络安全信息(告警)之间的相似度，进而决定新产生的告警的聚类归属。Alfonso Valdes通过手工定义的属性相似概率矩阵和函数来计算属性相似度，通过对多属性相似度做加权平均来计算网络安全信息间的整体相似度，算法如下。

$$Sim(X,Y)=\frac{\sum_j E_j SIM(X_j,Y_j)}{\sum_j E_j} \quad (1)$$

其中，X代表待匹配的告警，Y代表新告警，j代表告警属性的序号，E_j代表属性j的相似度期望，X_j, Y_j 代表在X和Y告警中其属性j的值。通过适当配置相似度标准参数，该方法可以分别分析网络安全信息间的冗余关系和关联关系。

基于概率聚类的网络安全信息分析技术，在相似性标准

基金项目：国家信息安全中心基金资助项目（2001-研2-A-005）
作者简介：彭雪娜(1979—)，女，博士生，主研方向：网络安全，网络管理；闻英友，博士；赵 宏，教授、博导
收稿日期：2006-02-24 **E-mail：**pengxn@neusoft.com

取值适当的情况下，能够较好地实现对网络安全事件信息的有效聚集和关联。但是由于相似性标准的取值是由用户手工调整的，系统并不理解不同取值以及分析结果之间的差别，因此，该技术在使用中，其使用效果严重依赖于使用者所掌握的安全知识。

2.2 基于模式识别的分析技术

基于模式识别的分析技术通过描述实际发生的网络安全事件的网络安全信息序列与预先定义的攻击场景模式相匹配，来识别网络安全信息之间的关系和网络中发生的实际攻击场景及目的。

基于模式识别的分析技术需要解决的核心和基础问题就是攻击场景的建模问题。目前常用的攻击场景建模手段包括：自动机，时序逻辑模型，专家系统以及对上述模型的扩展和演变等。目前，很多旨在分析网络安全信息间关系的研究都是基于模式识别技术来实现的。

Richard Kemmer 和 Giovanni Vigna 领导的 STAT 项目组提出了一种基于有限自动机建模的攻击场景识别技术。该技术用自动机的状态节点来描述攻击场景所可能处于的各种状态，用自动机的状态迁移来描述攻击场景中的攻击事件信息。分析引擎根据预定义的攻击场景对网络中产生的安全信息流进行匹配，进而识别相应的攻击场景。

Benjamin Morin和Herve Debar提出了一种基于时序模式识别的攻击场景识别技术[2]。所谓时序模式识别技术(Chronicle Recognition)是指按照一定的时序逻辑推理方法，基于一定的时序模式规则对输入的时序信息流进行匹配，从而将网络安全信息按照预定义的时序模式进行关联，并识别出符合时序模式的输入信息流。时序模式识别技术对自动机技术而言的优势在于它能够灵活地描述事件间的时序关系，并能够随着时间的推移自动地取消对一些不可能成立的时序模式的识别。nimda蠕虫的时序逻辑描述如下：

```
1 chronicle nimda[?source, ?target]
2 {
3 occurs((1,2), alarm[iis_code_red_ii_root_exe, ?source, ?target],
   (t, t+2000))
4 occurs((1,4), alarm[iis_decode_bug, ?source, ?target], (t, t+2000))
5 occurs((1,14), alarm[iis_cmd_exe, ?source, ?target], (t, t+2000))
6 occurs((1,3), alarm[web_dot_dot, ?source, ?target], (t, t+2000))
7 occurs((1,2), alarm[iis_unicode, ?source, ?target], (t, t+2000))
8 occurs((1,1), alarm[iis_unicode2, ?source, ?target], (t, t+2000))
9 occurs((1,1), alarm[iis_unicode3, ?source, ?target], (t, t+2000))
10 occurs((1,1), alarm[iis_decoee_bug3, ?source, ?target],
    (t, t+2000))
11 occurs((1,1), alarm[iis_decoee_bug2, ?source, ?target],
    (t, t+2000))
12 occurs((1,1), alarm[iis_decoee_bug4, ?source, ?target],
    (t, t+2000))
13
14 when recognized {
15 emit event(alarm[nimda, ?source, ?target], t);
16 }
17 }
```

Steven Cheung等提出了一种基于专家系统的攻击场景识别技术[3]。主要思想是将攻击场景描述为一系列模块，每个模块描述了一个攻击场景，攻击场景可以由一系列攻击子场景和攻击事件组成，当攻击子场景或事件满足一定约束条件时，则认为发生了特定的攻击场景，并将形成一个描述该攻击场景的抽象事件，其攻击模型如图 1 所示。

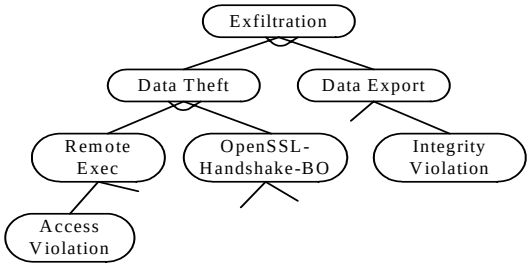
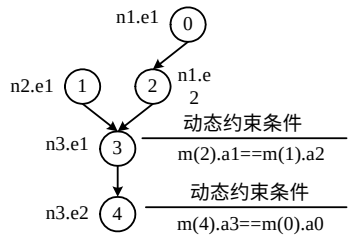


图 1　CAM 中的攻击模式

Christopher Kruegel和Thomas Toth提出了一种P2P的基于模式图识别的攻击场景识别技术[4]。Christopher Kruegel采用树形的攻击场景模式来实现对攻击场景的建模，如图 2。

```
ATTACK “ Sample ”
[n1,n2,n3]
n1{
  e1[$x=a0;]
  send(n3):e2[$y=a1;]
}
n2{
  send(n3):e1[a2==$y;]
}
n3{
  e1[]
  e2[a3==$x;]
}
```



n1.e2事件的发生创建了消息（2，发生时间，（a1,a1的值））

图 2　Quicksand 中模式图及节点约束条件

树形攻击场景模式图描述了位于不同主机上的事件序列，这些序列通过“ 发送 ”事件连接起来。攻击场景图的每一个节点上都描述了一个事件以及对该事件的约束，每台主机上的事件序列都是按照发生次序通过有向边来连接起来的，不同主机上的事件序列是通过由“ 发送 ”事件引出的有向边指向下一个主机上的第 1 个事件连接起来的。相关理论分析证明了这种分布式的基于攻击场景模式图匹配的方法具有良好的扩展性和容错性。

基于模式识别技术的分析能够分析出网络安全信息之间的关联关系，但是仍然存在以下 4 方面的问题： (1)由于该技术必须基于预先定义的攻击场景模式集，此类技术只能识别出已知的攻击场景，而无法识别未知的攻击场景。(2)虽然可以通过概括或增加攻击场景模式来扩大攻击场景识别的范围，但是由于可能的攻击场景模式繁多，因此该模型中解决对攻击场景的概括和穷举是困难的。(3)信息来自底层网络安全部件，当底层部件存在漏报问题时，上部不能保证对不完备输入信息所描述的攻击场景的准确识别。(4)从实际应用效果方面来讲，由于软件行为相对简单、确定、可预期，而人类攻击者的行为相对复杂、不确定和不可预期，因此这种技术只适合于识别软件攻击场景，而不太适合于识别人类攻击者的攻击场景。

2.3 基于网络安全信息因果关系的分析技术

基于网络安全信息因果关系的分析技术的基本思想是网络安全事件存在固有的因果关系，基于这种固有的因果关系可以将描述网络安全事件的网络安全信息很好地关联起来，进而形成可以描述网络安全事件之间关系的攻击场景图。

基于网络安全信息因果关系的分析技术的基本方法是为

各种安全相关的网络事件类型定义因果关联知识，并基于因果关联的算法，识别出输入网络安全信息流中安全信息之间的关联关系，最终形成攻击场景图。网络安全事件的因果关联知识定义为三元组<prerequisite, eventType, consequence>，其中 prerequisite 定义事件成功发生所需要的前提条件集合，包括网络或系统需要符合的一些事实以及攻击者所应具备的能力；consequence 定义事件成功发生后可能会产生的结果集合，包括攻击者对某些事实的发现或某种能力的获取；eventType 是事件的类型标识，知识实例如图 3 所示。

事件类型	前提	结论
Email_Debug	ExistService(DestIP, DestPort) && SendmailInDebugMode(DestIP)	{GainAccess(DestIP)}
Email_Ehlo	ExistService(DestIP, DestPort) && SMTPSupportEhlo(DestIP)	{GainSMTPInfo(SrcIP, DestIP)}
…	…	…
…	…	…
FTP_Pass	ExistService(DestIP, DestPort)	
FTP_Put	ExistService(DestIP, DestPort)	{SystemCompromised(DestIP)}

图 3 事件因果知识库

因果关联算法的基本思想是：A、B 两个事件(A 事件发生在 B 事件之前)能够关联，当且仅当 A 事件的结果集合中的元素能够完全或部分匹配 B 事件的前提条件集合中的元素。基于关联知识，采用关联算法可以识别出输入事件之间的关系，形成攻击场景图。攻击场景图都是一个有向图，其节点表示安全相关事件，包括告警事件或特殊应用事件；其节点间的有向边表示两个节点之间的关联关系。图 4 展示了一个关联后的攻击场景。

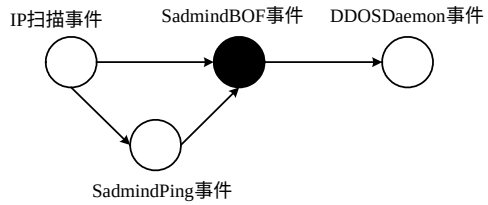


图 4 事件因果关联结果

Peng Ning所领导的TIAA项目[5]、ONERA的Frederic Cuppens所领导的MIRADOR项目[6]、UCDavis的Steven Templeton和Karl Levitt的JIGSAW项目[7]都是基于因果关联技术来实现的。这些项目都已经实现了相应的原型系统，其中TIAA项目和MIRADOR项目都对网络安全信息因果关联技术的理论研究作出了很大贡献。

因果关联技术虽然从本质上揭示了网络安全事件之间的联系，单纯的采用因果关联的技术仍然存在一定问题。采用这种方法关联后的攻击场景与实际攻击场景很可能存在一些差异，而这些差异会妨碍管理员对攻击场景的认识，甚至在一定程度上误导管理员。具体问题表现为两方面：(1)攻击场景的缺损问题。由于底层检测部件可能漏报响应信息，会导致攻击场景缺损，该问题是比较难于解决的。鉴于信息的缺乏，目前主要采用推理或假设的方法来试图解决。(2)攻击场景的高噪声问题。由于关联算法难于区分网络安全信息与攻击者的意愿之间的相关关系，因此降低攻击场景的噪声也是很困难的，目前通常采用限制某些高噪声信息关联的方法来降低攻击场景噪声。

2.4 基于网络安全信息统计因果关系的分析技术

基于网络安全信息统计因果关系的网络安全信息分析技术的基本思想是网络安全事件之间存在某种统计因果关系，这种统计因果关系能够把不同的网络安全信息关联起来，通过对当前网络安全信息流进行统计因果分析，可以识别出它们所描述的网络安全事件间的因果关系。

Xinzhou Qin和Wenke Lee提出了一种Granger Causality Test(GCT)的算法来对网络安全信息进行关联[8]。GCT算法通过用统计函数计算两个时序事件的Granger Causality Index(GCI)值，将该值与Fisher分布测试的特定值相比较，确定两个时序事件间的统计因果关系。GCT算法在网络安全信息的关联分析中的具体应用方法，是将网络安全信息流看作时序事件流，通过计算两个不同事件(X_i,Y)间的GCI，择优选取前m个X_i事件，判定X_i事件与Y事件的统计因果关系。

基于统计因果分析技术的网络安全信息关联具有很强的技术创新性，给网络安全信息关联技术的发展开拓了一个新的方向。但是由于统计因果分析技术本身发展得并不成熟，因此其有效性尚值得商榷。

3 结论与展望

网络安全信息分析技术主要解决的是在大量的、不准确的、重复的网络安全信息中识别出真正具有较高风险的网络安全事件并分析和选择出合适安全措施和响应策略，以降低高风险事件给网络带来的损失和影响。该技术近几年发展迅速，主要表现在关系分析和场景发现技术方面。除了本文中提到的技术中仍然存在的问题之外，在网络安全信息分析技术领域还有以下几个方向也亟待解决，包括：网络安全信息规范化标准建立，多种来源的网络安全信息的融合分析，网络安全事件的预测和预警，网络安全事件响应决策支持技术，网络安全管理和网络管理技术的融合等。这些问题的深入研究将对网络安全信息分析技术的发展起到至关重要的作用，也将是未来一段时间内国内外学术界在该领域的研究重点。

参考文献

1 Alfonso V, Keith S. Probabilistic Alert Correlation[C]. Proc. of the 4th International Symposium on Recent Advances in Intrusion Detection. Springer-Verlag, 2001.

2 Benjamin M, Herve D. Correlation of Intrusion Symptoms: An Application of Chronicles[C]. Proc. of the 6th International Symposium on Recent Advances in Intrusion Detection, Pittsburgh, PA. USA: Springer-Verleg, 2003.

3 Steven C, Ulf L, Martin F. Modeling Multistep Cyber Attacks for Scenario Recognition[C]. Proc. of Third DARPA Information Survivability Conference and Exposition, Washington, 2003.

4 Christopher K, Thomas T, Clemens K. Decentralized Event Correlation for Intrusion Detection[D]. Technical University of Vienna Information Systems Institute, 2002.

5 Peng N. Techniques and Tools for Analyzing Intrusion Alerts[J]. ACM Trans. on Inf. Syst. Secur., 2004, 7(2): 274-318.

6 Frederic C, Alexandre M. Alert Correlation in a Cooperative Intrusion Detection Framework[C]. Proc. of IEEE Symposium on Security and Privacy, Oakland, California, USA, 2002.

7 Steven T, Karl L. A Requires/provides Model for Computer Attacks[C]. Proc. of Workshop on New Security Paradigms. Ballycotton, County Cork, Ireland, 2000.

8 Qin Xinzhou , Lee Wenke. Statistical Causality Analysis of INFOSEC Alert Data[C]. Proc. of the 6th International Symposium on Recent Advances in Intrusion Detection, Pittsburgh, PA, 2003.