实习信息资料

公司简介：

北京北信源软件股份有限公司创立于 1996 年，注册资本 14.5 亿，2012 年于深交所创业板上市，股票代码：300352。北信源是中国信息安全龙头企业之一，历经 20 多年发展，从终端安全、数据安全向大数据安全、云安全、移动安全、物联网安全等方向全面拓展。北信源总部位于北京，下设多个全资子分公司及五大研发中心。拥有 1200 多名信息安全专业人员，构建了全国七大区、近三十个省市的营销与服务网络。

北信源公司多次荣获国务院颁发的国家科学技术进步二等奖等荣誉，在标准制定、重大专项等方面积极配合政府行动，同国家政府部门、国内著名院校、国际顶级 IT 厂商长期保持战略合作关系，并成功打造信息安全知名品牌“北信源 VRV”。北信源顺应市场趋势，启动并实施全新战略规划，布局信息安全、互联网、大数据为公司三大战略方向，致力将北信源逐步打造成为物联网时代下智慧安全的全面解决方案提供商。

在信息安全领域，构筑了三纵四横的新一代终端安全管理体系，从内网安全、数据安全、边界安全、防病毒等方向对 Windows 终端、国产终端、移动终端、虚拟化终端提供全方位、立体化的安全保护。产品覆盖政府、军队军工、公安、金融、能源等重要行业数千

家单位，成功部署数千万终端。

在互联网领域，全力打造的“安全通信聚合平台”—信源豆豆（Linkdood），打造跨终端、全方位、安全可信的通信聚合平台；为用户提供了即时通信、协同办公、应急指挥、任务管理、ERP 改造、应用开发、万物互联、互联互通等多层次的平台服务。

公司的大数据平台是企业级大数据处理、分析和挖掘平台，通过采集终端行为、网络流量和安全设备等数据，依托人工智能算法和深度学习引擎，对于用户行为和业务数据进行分析评估，将综合态势、威胁态势、攻击态势、人员态势、资产态势、应用态势、流量态势、脆弱性态势这八大态势进行集中处理和可视化处理，帮助用户主动应对威胁和风险，时刻掌握全网安全态势和业务状况，为国家关键信息基础设施和重要信息系统提供了重要的安全保障。

北信源将秉承“信息之源、信誉之源、信心之源”的核心理念，依靠领先的产品、技术与服务，本着一如既往继往开来的创新姿态，为构筑自主可控的信息网络安全体系而不懈努力。

公司全称：

北京北信源软件股份有限公司

联系地址：

北京总部

地址：北京市海淀区中关村南大街 34 号中关村科技发展大厦 C 座 16 层

邮编：100081

电话：010-62140485/86/87

传真：010-62140468

职位名称：

研发工程师（实习）

签字负责人：

杨泳

研发部部门总经理

实习主管：

汤凌峰

内网和工控安全部门经理

接待人员：

刘泽欣

研发助理

项目名称：工控安全中数据流分析器/管理平台

数据流分析器/管理平台

目　录

第一章　引言

1.1 设计背景

现代工控自动化要求全过程的实时监控、实时数据的处理、工控安全以及生产信息的共享。工业控制系统的流量大多具有简洁性和周期性的规律，基于以上两种规律设计了一个数据流分析器。主要功能包括：可视化人机操作界面、工厂的数据接收和过滤、数据分析、威胁性检测等。

1.2 设计意义

工控安全关系着国家安全，从工控网络流量中发现威胁行为是保护工控系统的手段之一。因此，针对工业控制系统流量的特点开发有效的数据分析和威胁检测手段

第二章　术语解释

2.1 通讯规则

截取数据包五元组数据，提取 MAC 对、PORT 对、IP 对、协议类型形成通讯记录，对收到的数据块处理，重复的数据保留第一行，得到通讯规则。

2.2 杂包

按通讯规则库的数据结构对数据包逐项比较，遍历通讯规则数据库后，无匹配的称为杂包。

2.3 非杂包

按通讯规则库的数据结构对数据包逐项比较，遍历通讯规则数据库后，有匹配的称为非杂包。

2.1 通讯规则：针对要处理数据包的一个拦截操作并提取出相关信息的操作。

子术语：①数据包：包(Packet)是TCP/IP协议通信传输中的数据单位，一般也称"数据包"。(与Web对应)

②五元组数据：指源IP，源端口(Port)，目的IP地址，目的端口和传输层协议。（R，U，D，dom，F）

③通讯记录：对拦截的五元组数据提取，其中 Mac对，Port对，IP对，协议类型 组成通讯记录。

注：删除冗余重复数据

④数据块：数据块是一组或几组按顺序连续排列在一起的记录，是主存储器与输入设备、输出设备或外存储器之间进行传输的数据单位。(与BDD对应)

2.2 杂包：根据规则筛选出的无匹配数据包。

①通讯规则库：由2.1组成的集合体

②

2.3 非杂包：与2.2相反。

第三章　数据块

3.1 威胁载荷库

从外部导入威胁数据，导出尚未匹配的数据

3.2 杂包数据库

杂包数据包加时间戳，数据包保存到杂包数据库

3.3 非杂包包头库

非杂包包头数据包加时间戳，保存到非杂包包头库

3.4 分析数据库：

包括网元数据表、协议流量表、协议曲线数据表、IP 对流量表、协议 IP 流量分布曲线数据表。

3.5 用户及日志数据库

包括管理员、用户管理、用户登录、操作日志等

3.6 通讯规则库

读入非杂包通讯关系缓存数据，清空缓存区，

3.1 威胁载荷库：数据库之威胁数据

3.2 杂包DDD：对杂包加个时间字符序列，然后保存到一个数据库里

① 时间戳 (timestamp)：一个能表示一份数据在某个特定时间之前已经存在的、完整的、可验证的数据，通常是一个字符序列，唯一地标识某一刻的时间。

3.3 非杂包包头库：对非杂包加时间戳，保存至一个DDD里

3.4 分析数据库：以下5种表的结合体

① 网元数据表：由最小单位的网络元素（网元）组成的一张数据表（Table）

1.1 网元 (Network element)：网管系统中的网元其实和这个差不多，简单理解就是网络中的元素，网络中的设备。总之，网元是网络管理中可以监视和管理的最小单位，值得注意的是，网络元素与网元和被管设备是同义语，但被管设备容易被人误解成硬件。

② 协议流量表

2.1 流量表：流量指被动的数据表

③ 协议曲线流量表：图形化①

④ IP对流量表：针对IP对的流量（流入+流出）变化表

⑤ 协议IP流量分布曲线：图形化④

5.1 协议IP

3.5 用户及日志DDD：缓存等

3.6 通讯规则库：通讯规则不止一个，放到DDD中统一管理，

注：完成导入时清空缓存。

第四章　程序模块

4.1　交换机控制

4.1.1 人机界面

用户在人机界面选择交换机过滤启动和停止。

4.1.2 交换机模式控制

读取通讯规则并发送通讯规则给交换机。

4.1.3 交换机控制

采用 SSH 协议连接交换机，完成通用配置。

接收通讯规则后发送通讯规则给交换机，发送过滤启动（或停止）给交换机。

4.1.4 交换机

完成 SSH 连接服务、通用配置、数据镜像配置，接收通讯规则，将数据转换为 ACL。开始过滤。

访问控制列表：访问控制列表(ACL)是一种基于包过滤的访问控制技术，它可以根据设定的条件对接口上的数据包进行过滤，允许其通过或丢弃。访问控制列表被广泛地应用于路由器和三层交换机，借助于访问控制列表，可以有效地控制用户对网络的访问，从而最大程度地保障网络安全。

4.2 数据接收

4.2.1 网卡抓包

从指定网卡中抓包。

4.2.2 过滤 ARP 包

判断是否为 ARP 包，是则丢弃，并对 ARP 包计数，若不是 ARP 包则进行下一步过滤。

4.2.3 过滤广播包

从队列中逐包读入。读取五元组数据，提取 MAC。若 MAC 为 0.0.0.255，则丢弃该包；若不是，则加入队列 4。

4.2.4 数据接收准备

建立三个队列 A、B、C 初始化全为 0。队列 B 接收数据包，排队，先入先出；队列 A 缓存数据包，先入先出。每个数据包在队列 A 中缓存 100ms，超时数据包移入队列 B。队列 C 接收网卡数据包。

4.2.5 过滤重复包（会出现两倍的数据包）

队列 A 中每个数据包缓存 100ms，移入队列 B；队列 C 的后续数据包 X 与队列 A 的数据包逐包 XOR 操作，若 XOR=1 则将数据包 X 加入队列 A，否则将数据包 X 丢弃。

4.3. 规则学习

4.3.1 规则学习

从队列 B 逐包读入，截取数据包五元组数据，提取 MAC 对、IP 对、PORT

对、协议类型形成通讯记录。发送提取的数据块进行数据处理。

4.3.2 数据处理

对收到的数据块进行处理，行重复数据保留一行。列 1 至列 4 中重复数据保留第一行，其他行填 0，列 5 填 1。写入通讯规则库数据库。

4.3.3 通讯规则导入/导出

将通讯规则导入到文件；从电子表格导入到通讯数据库。

4.3.4 更新控制（待修改）

Q=0，则跳转至 end；Q=1，读入通讯规则数据写入缓冲区
通讯规则发给 Z，同时写入通讯规则数据缓冲区
置 Q=0，end

4.4 数据过滤

检测标志位，从队列 B 读入数据包，应用通讯规则过滤数据包。过滤规则：按通讯规则库数据结构逐项比较，相同项行不变，列加一；不同项行加一，列归一。遍历通讯规则后，无匹配项认定为杂包，提取杂包通讯关系；非杂包提取流包头和计数，加时间戳。输出杂包报警信息到用户界面。

读取杂包数据库和非杂包数据库的 MAC/IP 字段分别存储，数据有更新则读取。写入网元数据表。

4.5 威胁处理

4.5.1 杂包载荷威胁分析

杂包数据 MD5 与威胁数据库对比，匹配到内容提示，无法匹配的存入威胁载荷库，等待导出人工分析，展示威胁分析。

4.5.2 威胁态势输出

输出本系统存在的威胁信息、杂包比例，系统名称、威胁名称、活动状态和时间段。

4.6 数据分析

4.6.1 网元显示

读取杂包数据库和非杂包数据库的 MAC/IP 字段分别存储，数据有更新则读取，写入网元数据表。

4.6.2 按协议展示流量

读取非杂包数据，以日为周期，按协议计算分钟包/字节流量、5 分钟包/字节流量，存入协议流量表中；以分钟（秒）为步长，计算各个协议的包/字节数据，按日周期计入协议曲线记录表。

杂包同上处理。

4.6.3 按协议流量分布

计算各协议按 IP 分布的流量,读取非杂包数据。以日为周期，按 IP 计算分钟包/字节流量、5 分钟包/字节流量，存入 IP 对流量表中；以分钟（秒）为步长，计算各 IP 对上的包/字节数据，按日周期计入协议 IP 流量分布曲线记录表。

杂包同上处理。

4.7 流量展示

4.7.1 网元显示

读取网元数据包送入网元显示界面，编辑界面。

4.7.2 按协议展示流量

读取协议流量表的数据，按协议列表展示。

4.7.3 按协议流量曲线展示

读取协议流量曲线数据表，按协议展示曲线。也可选择不同协议展示。

4.7.4 协议流量分布展示

读取 IP 对流量表，按 IP 对列表展示流量数据，点击详情进入曲线展示。

4.7.5 协议 IP 流量曲线展示

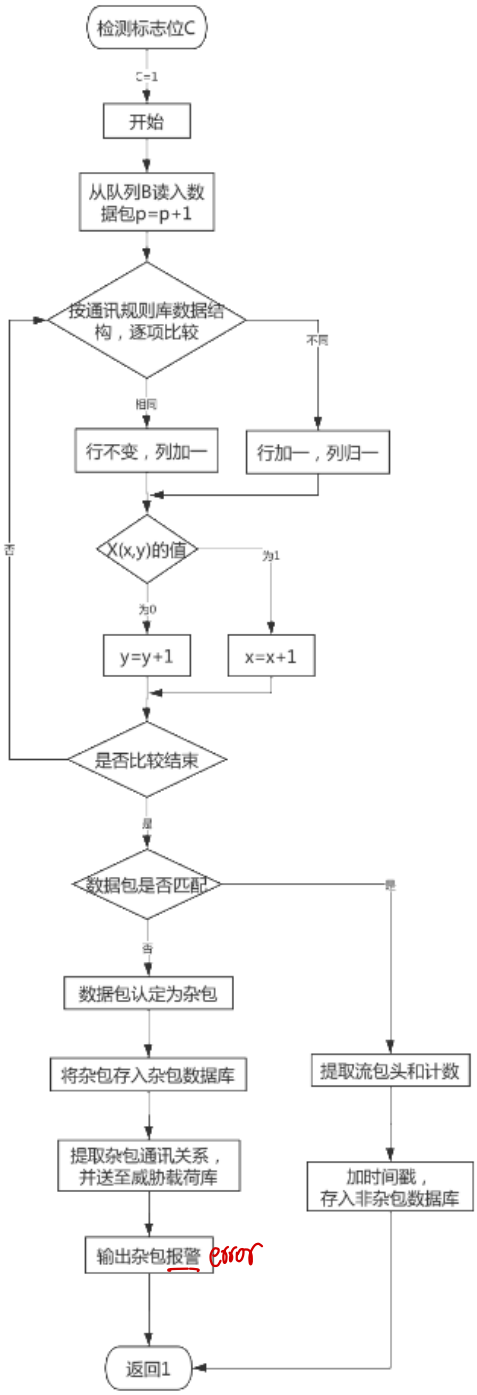
读取协议 IP 流量分布曲线数据表，按 IP 对分布展示协议流量曲线。

4.7.6 按时间戳展示包分布

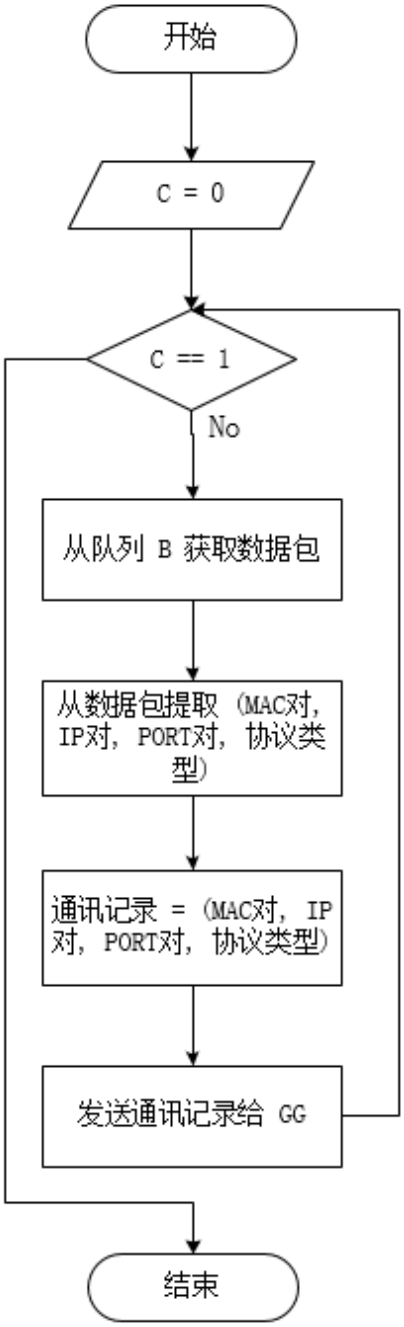
读取杂包/非杂包数据库，按时间戳展示数据包分布。

第五章　流程图

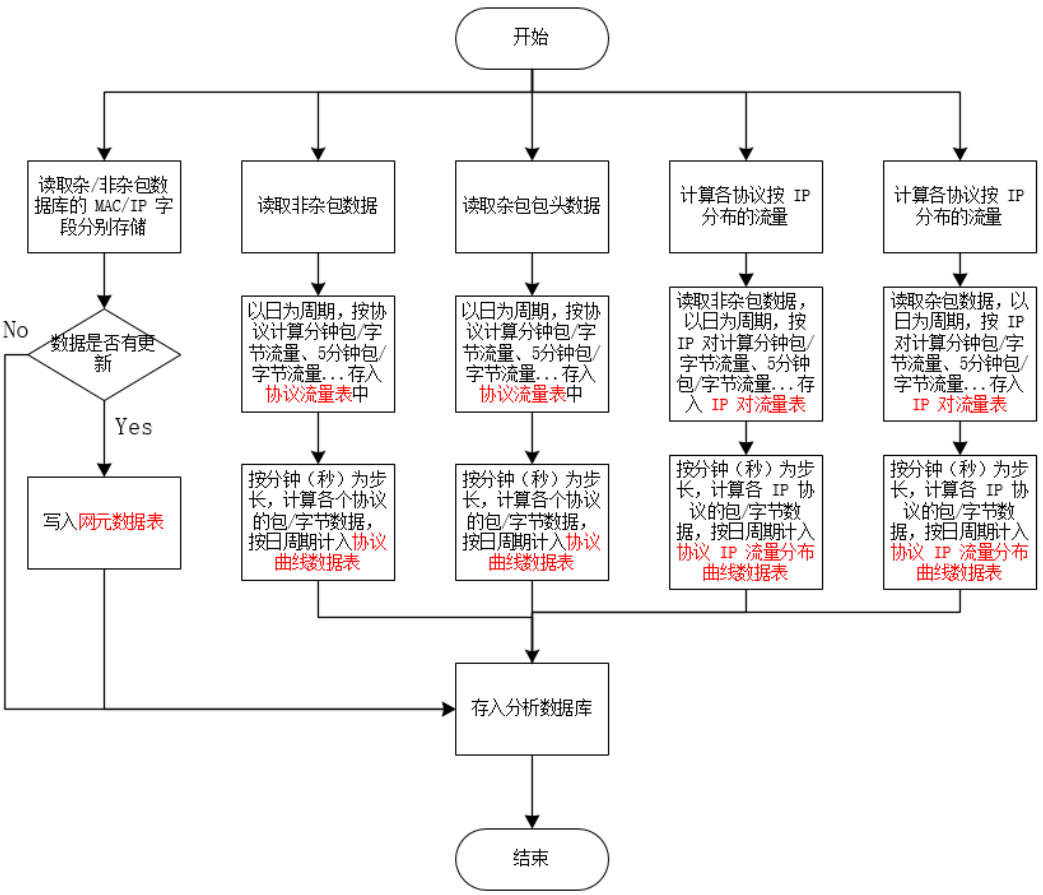
5.1 抓包模块



5.2 学习模块


开始
C = 0
C == 1
No
从队列 B 获取数据包
从数据包提取（MAC对，IP对，PORT对，协议类型）
通讯记录 = （MAC对，IP对，PORT对，协议类型）
发送通讯记录给 GG
结束

5.3 分析功能模块


开始
读取杂/非杂包数据库的 MAC/IP 字段分别存储
读取非杂包数据
读取杂包包头数据
计算各协议按 IP 分布的流量
计算各协议按 IP 分布的流量
No
数据是否有更新
Yes
以日为周期，按协议计算分钟包/字节流量、5分钟包/字节流量...存入协议流量表中
以日为周期，按协议计算分钟包/字节流量、5分钟包/字节流量...存入协议流量表中
读取非杂包数据，以日为周期，按 IP 对计算分钟包/字节流量、5分钟包/字节流量...存入 IP 对流量表
读取杂包数据，以日为周期，按 IP 对计算分钟包/字节流量、5分钟包/字节流量...存入 IP 对流量表
写入网元数据表
按分钟（秒）为步长，计算各个协议的包/字节数据，按日周期计入协议曲线数据表
按分钟（秒）为步长，计算各个协议的包/字节数据，按日周期计入协议曲线数据表
按分钟（秒）为步长，计算各 IP 协议的包/字节数据，按日周期计入协议 IP 流量分布曲线数据表
按分钟（秒）为步长，计算各 IP 协议的包/字节数据，按日周期计入协议 IP 流量分布曲线数据表
存入分析数据库
结束