Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2531410 A1 2007/06/23

(21) 2 531 410

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(22) Date de dépôt/Filing Date: 2005/12/23

(41) Mise à la disp. pub./Open to Public Insp.: 2007/06/23

(51) Cl.Int./Int.Cl. *H04L 12/26* (2006.01),
H04L 29/14 (2006.01), *H04L 9/00* (2006.01)

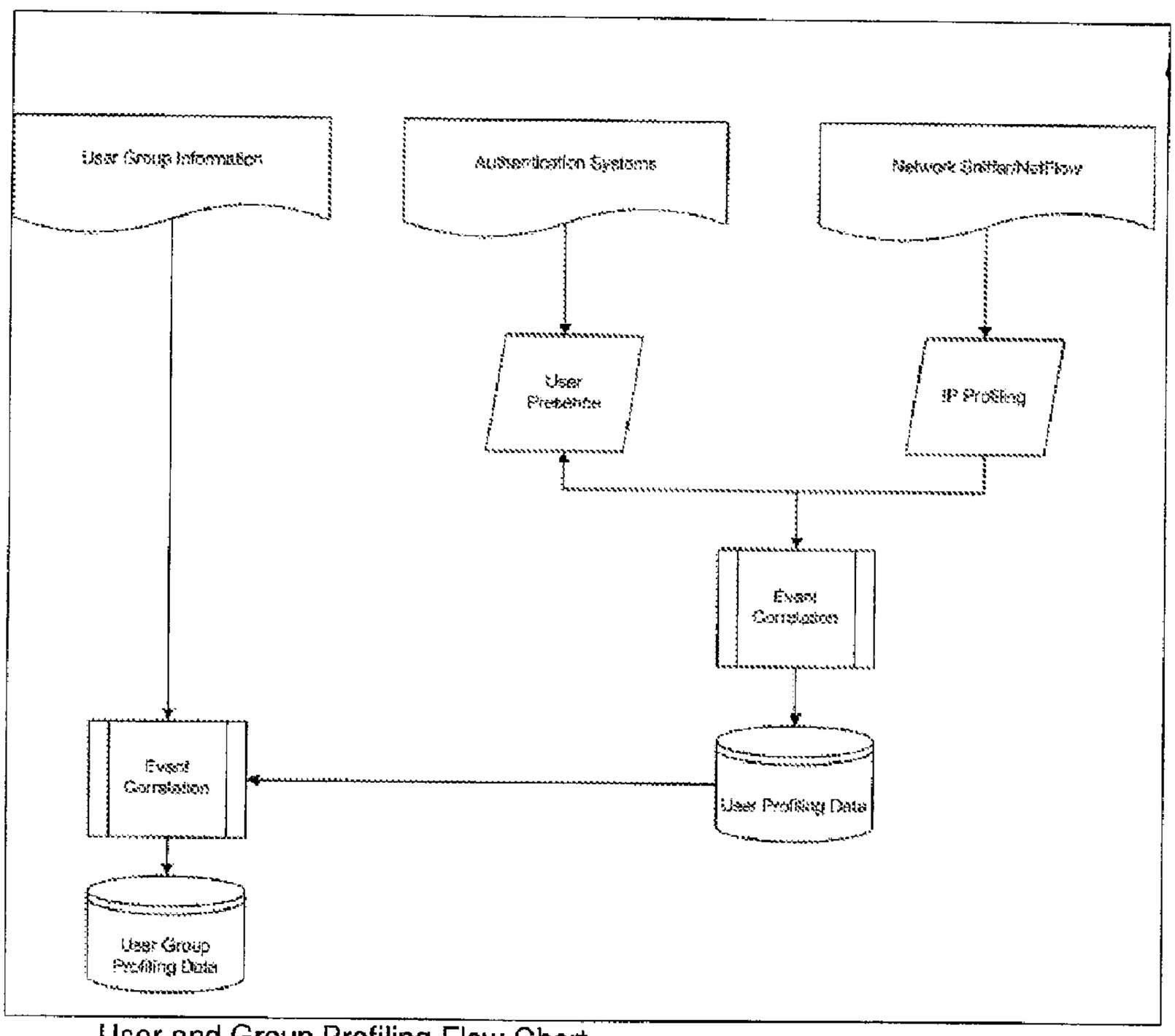
(71) Demandeur/Applicant:
SNIPE NETWORK SECURITY CORPORATION, CA

(72) Inventeurs/Inventors:
LIN, XIAODONG, CA;
YONG, YUH MING (PETER), CA

(74) Agent: BERESKIN & PARR

(54) Titre : DETECTION DES ANOMALIES D'UN RESEAU, BASEE SUR LE COMPORTEMENT, EN FONCTION DU PROFILAGE D'UN UTILISATEUR ET D'UN GROUPE D'UTILISATEURS

(54) Title: BEHAVIOURAL-BASED NETWORK ANOMALY DETECTION BASED ON USER AND GROUP PROFILING



User and Group Profiling Flow Chart.

(57) Abrégé/Abstract:
A baseline can be defined using specific attributes of the network traffic. Using the established baseline, deviation can then be measured to detect anomaly on the network. The accuracy of the baseline is the most important criterion of any effective network

CA 2531410 A1 2007/06/23

(21) 2 531 410

(13) A1

(57) Abrégé(suite)/Abstract(continued):
anomaly detection technique. In a local area network (LAN) environment, the attributes change very frequently by many change agents; for example, new entities, such as users, application, and network-enabled devices, added to and removed from the LAN environment. The invention provides an improved method of establishing a baseline for network anomaly detection based on user's behaviour profiling. A user behaviour profiling is a distinct network usage pattern pertaining to a specific individual user operating on the LAN environment. No two users profiling would be the same. A group of users that have similar network usage attributes can be extrapolated using data mining technique to establish a group profiling baseline to detect network usage anomaly. By combining user and group profiling, a network anomaly detection system can measure subtle shift in network usage and as a result separate good user's network usage behaviour from the bad one. Using the said technique, a lower rate of false positives of network anomaly can be created that is suitable to operate in a highly dynamic LAN environment.

ABSTRACT:

A baseline can be defined using specific attributes of the network traffic. Using the established baseline, deviation can then be measured to detect anomaly on the network. The accuracy of the baseline is the most important criterion of any effective network anomaly detection technique. In a local area network (LAN) environment, the attributes change very frequently by many change agents; for example, new entities, such as users, application, and network-enabled devices, added to and removed from the LAN environment. The invention provides an improved method of establishing a baseline for network anomaly detection based on user's behaviour profiling. A user behaviour profiling is a distinct network usage pattern pertaining to a specific individual user operating on the LAN environment. No two users profiling would be the same. A group of users that have similar network usage attributes can be extrapolated using data mining technique to establish a group profiling baseline to detect network usage anomaly. By combining user and group profiling, a network anomaly detection system can measure subtle shift in network usage and as a result separate good user's network usage behaviour from the bad one. Using the said technique, a lower rate of false positives of network anomaly can be created that is suitable to operate in a highly dynamic LAN environment.

BEHAVIOURAL-BASED NETWORK ANOMALY DETECTION BASED ON USER AND GROUP PROFILING

Field of the Invention

A baseline can be defined using specific attributes of a network traffic. Using the established baseline, deviation can then be measured to detect anomaly on the network. The accuracy of the baseline is the most important criterion of any effective network anomaly detection technique. In a local area network (LAN) environment, attributes change very frequently by many change agents; for example, new entities, such as users, application, and network-enabled devices, added to and removed from the LAN environment. The invention provides an improved method of establishing a baseline for network anomaly detection based on user's behaviour profiling. A user behaviour profiling is a distinct network usage pattern pertaining to a specific individual user operating on the LAN environment. No two users profiling would be the same. A group of users that have similar network usage attributes can be extrapolated using data mining technique to establish a group profiling baseline to detect network usage anomaly. By combining user and group profiling, a network anomaly detection system can measure subtle shifts in network usage and as a result separate good user's network usage behaviour from a bad one. Using this technique, a lower rate of false positives of network anomalies can be created that are suitable to operate in a highly dynamic LAN environment.

Background of the Invention/Description of the Prior Art

The topic on the anomaly based intrusion detection has been extensively studied in the past decade and witnessed so many security breaches made headlines. In order to improve weaknesses of signature based intrusion detection system (IDS), the anomaly detection systems come into play since in 1987 when Dorothy Denning presented a model of how an anomaly detection system could be implemented. The anomaly detection systems fall into six major categories, depending upon the methods they use to learn baseline behaviours and identify deviations from those established baselines. The six main detection types include neural networks, statistical analysis, signal processing, graph, payload and protocol-based systems.

Neural networks: Neural Networks have been proposed as a means of performing anomaly detection. Neural networks can be divided into two main algorithm types: those that employ supervised training algorithms, where in the learning phase, the network learns the desired output for a given input or pattern, and unsupervised training algorithms, where in the learning phase, the network learns without specifying the desired output. Research is being conducted regarding the application of neural network pattern recognition abilities to network behaviour anomaly detection but there are no commercial applications as of yet.

Statistical analysis: Network data points can be modeled using a stochastic distribution of any network traffic features, such as IP addresses or network ports. A baseline can be established by calculating the characteristics of the modeled network traffic feature distributions. Once a baseline is established, specific data points can be determined to be anomalies depending on their relationship to this established baseline. The major problem with statistical anomaly detection models is that as the number of variables or dimensions increases, the more difficult it becomes to accurately estimate distributions.

Signal processing based anomaly detection: Any network traffic feature can be modeled as a time series. A network anomaly would therefore be identified as correlated abrupt changes in network data. An abrupt change is defined as any change in the parameters of a time series that occurs on the order of the sampling period of the measurement of any of the chosen network traffic features. Whenever the change is large, this method produces similar results to the traditional threshold based statistical analysis method. However, the signal processing based method is also very effective in detecting minute changes which often occur at the early stage of an attack, such as Internet worm outbreak or server failure, and thus is extremely useful in reducing exploitation costs.

Graph based anomaly detection: It has been proposed that network anomalies can also be detected by graphing network connections. In such a graph, nodes represent network hosts and edges represent connections between these hosts. By observing how these graphs change over time, many types of anomalous usage can be detected. Examples include a particular host that does not usually connect to many machines suddenly establishes connections to several hosts it has never contacted before and may indicate that a machine has been compromised. Similarly, an activity such as a machine that has only ever connected to email and Web servers that begins

connecting to database servers would also be detected. Internet worms can be detected because of the way they spread. It would be unusual for a host to contact another host and shortly later both hosts begin contacting many other hosts, constantly perpetuating and enlarging this behaviour. The resulting graph can then be used to identify the source and propagation of the worm.

Protocol anomaly detection: Instead of training models on normal behaviour, protocol anomaly detectors build models of TCP/IP protocols using pre-built specifications. Since protocols are well defined, a normal use model can be created with greater accuracy and ease. Protocols are created with specifications, known as RFCs, to dictate proper use and communication. All connection oriented protocols have state meaning that certain events must take place at certain times. As a result, many protocol anomaly detectors are built as state machines. Each state corresponds to a part of the TCP connection, such as a server waiting for a response from a client. The transitions between the states describe the allowed and expected changes between states. When unexpected state changes occur, the model flags these changes as anomalous events.

Payload-based Anomaly Detection: Payload based anomaly detection is the method to detect anomalies. Payload based anomaly detection analyzes the bytes that are being transferred in the payloads of packets and looks for any anomalies in a payload packet's inherent structure. Generally, each application layer protocol will have its own unique structure that can be used to identify the protocol. By analyzing all traffic going to a particular port, for example Port 80, it can be determined if there is anything other than HTTP traffic travelling on that port. This is a necessary security precaution as firewalls generally admit all traffic on port 80 without any inspection of packet contents. Since any service can be configured to run on any prot, payload-based anomaly detection can protect against rogue port uses.

Network anomaly detection systems usually have a high rate of false positives. The reason is that the current network behaviour anomaly detection systems solely model network traffic. In reality, network traffic patterns, especially in LAN environment, are very dynamic and change frequently, which result in high rate of false positives.

One design consideration is that the LAN environment is highly dynamic and any number of things can change the network traffic patterns; for example, adding new services, adding new employees or adding new resources. Another design consideration is that network user habits are deterministic and once engrained, these habits are difficult to change.

Accordingly a more accurate and effective network anomaly detection system should be based on user behavioural profiling and assume the network environment is always dynamic and not static. These two attributes (i.e. dynamic LAN environment and deterministic human habits) are used to design a system that measures anomaly in a LAN environment. The new system can detect obvious and subtle network usage changes and therefore increase the accuracy of network anomaly detection and lower the rate of false positive alerts. The system uses user and group profiling to establish a baseline for comparison to detect network anomaly. User profiling reflects the user's normal behaviour (for example, the network resources used, and Web sites visited). A group profiling reflects a group of network users who have similar responsibilities or attributes (for example, a group of users who use certain types of network services). The system establishes a baseline for modelling user's behaviour on the LAN. The baseline is a representation of accepted user's behaviour on the network that is learned by the system over a period of time. The baseline can be learned by the system or explicitly specified by the network administrators, or both. Deviations from the baseline are analyzed for significance to identify anomalous network user's behaviour. This invention implements user and group behaviour anomaly detection to catch network anomalies such as unauthorized access, network abuse and misuse, unauthorized transmission of information to external network, and slow-moving and fast-moving worms and viruses.

Summary of the Invention

The new system in this invention deals with the complexity of LAN environment and network user's behaviour. The solution models these two attributes (i.e. dynamic LAN environment and complex network user's behaviour) to detect unknown and new network anomalies. Instead of modelling network traffic, the system focuses on modelling user's behaviour and building user and group profiling based on what the network users have done on the LAN. The system applies user profiling to reflect the user's normal behaviour, such as

the network services they used and the Web sites that they visited. Additionally, a group profiling for a group of users, who have similar responsibilities or attributes, can be established to reflect the common behaviour of majority members in the group that are considered good network usage behaviour based on the assumption that violators are just minority network users on the LAN.

It can be assumed that the network users on the LAN must have been authenticated before allowed access on the LAN or use any network services. Based on this assumption, the new system can trace the presence of network users on the LAN by interrogating the authentication server or installing a software agent in user's host machine to gather such information. The user presence information is then correlated with the network IP address that is used by the network user.

By correlating user presence and network information, a behavioural profiling can be established that uniquely reflects an individual user's distinct network usage and network traffic patterns. The distinct attributes of a specific user establish a baseline that is subsequently used to measure deviation. A set of users who have similar responsibilities or attributes can be specified as a group profiling by the system administrator and using the group profiling to establish a baseline that separates collective good and bad user's behaviour on the LAN environment.

By aggregating a set of user profilings, a group profiling can then be defined. The group profiling models a collection of users that exhibit similar and common behaviour patterns. The group profiling is used to detect subtle deviation from an individual user's normal behaviour on the LAN.

Brief Description of the Drawings

In the drawings, which form a part of this invention,

FIG. 1 is a flow chart of User and Group Profiling;

FIG. 2 is a sample user profiling raw data;

FIG. 3 is an illustration of user profiling of network services visited.

Detailed Description of the Invention

The new system is preferably composed of the following three components:

1. Learning user and group profiling - this is used to build user and group profiling database based on the information collected from network access authentication system and network devices, such as network switch and network tap.
2. Detection Engine - this is used to identify deviations from the established user and group profiling data (i.e. baseline or normal behaviour). These deviations are analyzed for significance and are then categorized as anomalous behaviour.
3. Graphic User Interface - this is used to monitor events and alerts and manage the detection engine by the network administrators.

The flow chart, as shown in FIG. 1, describes how the said system creates user and group profiling.

The system assumes that the network user has been authenticated before allowed access to the LAN and to use network services. In the case of the Microsoft Windows authentication scheme, there could be at least one domain controller that allows or denies access to network resources on the domain. Because the domain controller stores user authentication information, performs authentication, and enforces security policy for a Microsoft Windows domain, the new system would integrate with the Microsoft Windows domain controller to read the Microsoft Windows domain controller's log and fetch authentication log. The log files are then correlated to derive user's presence information that consists of user's log-in name, network IP address, and asset's network MAC address.

The new system can also use a software agent that is installed on the user's host machine to derive user's presence information.

Given the user's presence information, the system can obtain the network packets through various methods to build the user and group profiling. Some of the methods are (1) proprietary network packets collection protocol such as NetFlow, sFlow, and cFlow, (2) passive network TAP, and (3) SPAN port. The raw user profiling data, as illustrated in FIG. 2, would reveal information of the user's network activities - such as network services visited, type of services used, and method of network access.

If the user profiling raw data of a particular user is represented in the form of a histogram, the X-axis represents the network services visited and the Y-axis represents the number of network packets generated using the network services. Using the histogram as a probability distribution, the new system calculates the "entropy" (which is defined as a measurement of the degree of dispersion of a distribution) to evaluate any shifts in user behaviour. An entropy is calculated for each network service consumed by the user.

All entropies are normalized to provide a faster evaluation of anomalous score and to decide whether or not there are behavioural anomalies by comparing against the established baseline. Furthermore, those measurements, entropy of user visiting network services, could become input of any machine learning algorithms, such as ANN (Artificial Neural Network), SVM (Support Vector Machines), and create a detection engine and increase the accuracy of anomaly intrusion detection.

One use embodiment is in some port scanning techniques, which does not incur significant network traffic changes. In this use case, traditional network behaviour anomaly detection systems will not detect the exploit because the resultant change in network traffic may not be substantial enough to trigger a large deviation from the established baseline. However, suspicious and rare visiting network service of a user could incur significant deviation of his user profiling visiting service distribution, which results in immediate detection of this incident.

Another use embodiment is where a new employee is added to the internal network environment. In this case the network traffic baseline will be shifted and can cause a network behaviour anomaly detection system to generate many false positives. However, the system's user behavioural-based anomaly detection model is able to determine that there is a new user joining the network system and will not inaccurately flag this event as an alert.

A still further embodiment is where a new network application is added to the LAN. In this use case, traditional network behaviour anomaly detection systems could flag the event as a Trojan Horse attack. However, the new system model would detect the newly added network application in a passive way, and observe the change in user behaviour. Furthermore, the system would also detect the shift in group profiling behaviour. By correlating user and group behavioural shifts, a low level notice will be issued rather than the high level alert generated by the detection of a Trojan Horse.

By applying user behavioural anomaly detection techniques, the system could detect fast-moving and slow-moving network anomalies that manifest in a LAN environment whose network traffic is highly dynamic and the operating attributes change frequently.

WHAT IS CLAIMED IS:

1. In a LAN environment system wherein network traffic is highly dynamic and operating attributes change frequently, system means for applying profiling of user's network behaviour to define a baseline that is subsequently used to detect anomalous network usage and malicious network behaviour.

2. The system recited in claim 1, wherein user presence is correlated with network usage information to link an identity of a network user to his/her network usage patterns, the user's presence information including user's login information, a network IP address assigned to the user's host machine, and the user's host machine's network MAC address, the network usage information including IP address of network service, network protocol, entry point of network service, and type of network service.

3. The system recited in claim 2 desired from an authentication system that allows or denies network access and maintains a database of user authentication, data the authentication system taken from the group of Unix, Microsoft Windows domain controller and active direction, RADIUS, Microsoft Network Access Protection (NAP), Cisco Network Admission Control (NAC), 802.1x, and authentication systems that exhibit attributes of network access control and authentication data management.

4. The system recited in claim 2 obtained by sniffing network packets via passive network Tap device, SPAN port of managed switches, and NetFlow, sFlow, and cFlow data of vendor-specific network devices.

5. A collection of the user profilings from the system as recited in claim 2 which collective defines a group profiling, the group profiling consisting of a set of users who exhibit similar operating attributes in the LAN environment, the attributes being categorized by the user's roles and responsibilities in an organization.

6. The collection of user profilings as recited in claim 5 wherein the user's are employees in an R&D organization.

7. The collection of user profilings as recited in claim 5 wherein the users are defined by system administrators or imported from an authentication system, such as a Windows domain controllers.

8. The collection of user profilings as recited in claim 5 wherein the collection is used to establish a baseline of common behaviour of the group of users, the baseline being derived using data mining technique and then used to detect network usage anomalies, the group profiling representing normalized good behaviour of the group of users based on an assumption that a majority of members in the group exhibit good network usage behaviour.

9. The collection of user profilings as recited in claim 5 used to reduce an effect of baseline shift due to behaviour changes by a small subset of users within the group, the group profilings reflecting the common behaviour of majority users in the group, which are considered as good behaviour on the assumption that violators are minority users in the LAN environment and the majority of the users have normal acceptable network behaviour.

10. The system as recited in claim 2 wherein, when a user's network behaviour changes and deviations are too far off from the individual's user profiling baseline but similar deviations are also exhibited in other users in the same group, then the anomalies will be fedback to the system as newly discovered normal user behaviour, to re-establish user and group profiling baselines.

11. The system as recited in claim 10 wherein the detected collective shift in network behaviour establishes new user and group baselines and correlates to similar changes in behaviour of a majority users in the same group profiling.

12. The system as recited in claim 10 wherein the changes in behaviour attributed to the majority user is appended into the user and group profilings.

13. The system as recited in claim 10 wherein the user and group profilings are used to monitor normal network usage and allow security policy to be enforced at the user level.

Application number/numéro de demande: 2531410

Figures: 2-3

Pages:

DKW-IP

Unscannable items
received with this application
(Request original documents in File Prep. Section on the 10th Floor)

Documents reçus avec cette demande ne pouvant être balayés
(Commander les documents originaux dans la section de préparation des dossiers au 10ième étage)

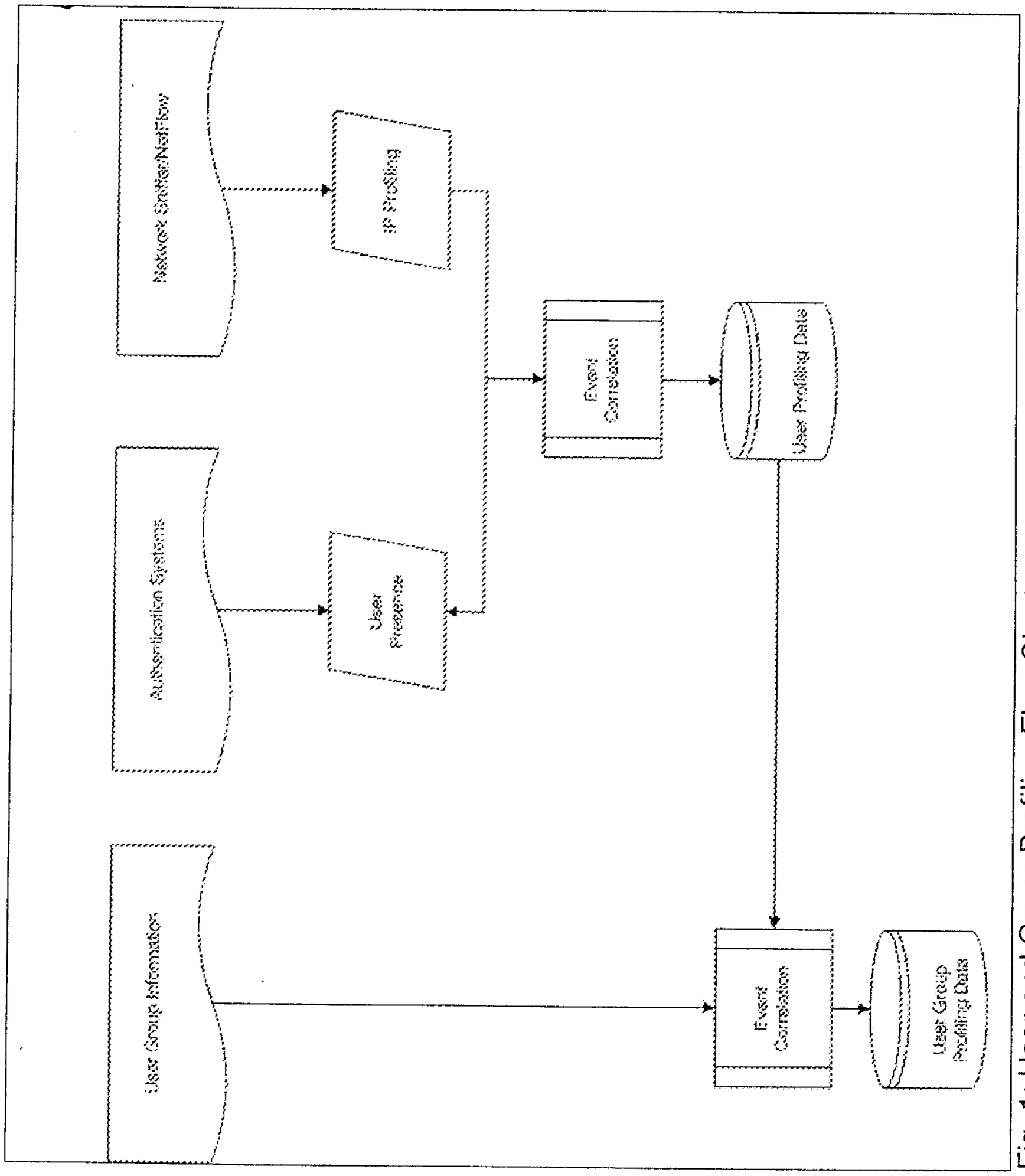
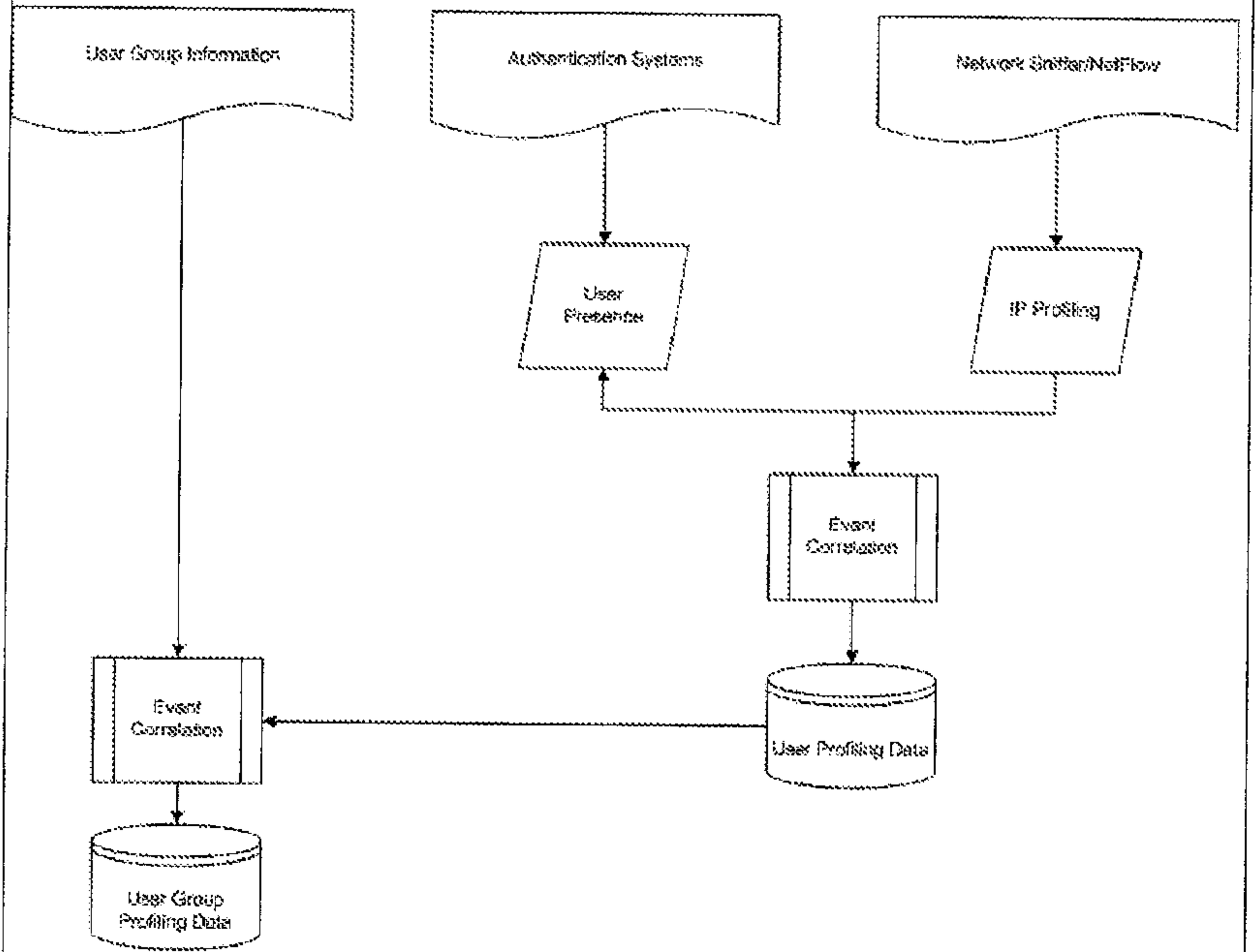


Fig. 1: User and Group Profiling Flow Chart.

Sudlaysod & Singlehurst
PATENT AGENTS



User and Group Profiling Flow Chart.