2013 IEEE Security and Privacy Workshops

Differentiating User Authentication Graphs

Alexander D. Kent*, Lorie M. Liebrock†
*Los Alamos National Laboratory
alex@lanl.gov
†New Mexico Institute of Mining and Technology
liebrock@nmt.edu

***Abstract*—Authentication using centralized methods is a primary trust mechanism within most large-scale, enterprise computer networks. This paper proposes using graphs to represent user authentication activity within the network. Using this mechanism over a real enterprise network dataset, we find that non-privileged users and users with system administration privileges have distinguishable graph attributes in terms of size and complexity. In addition, we find that user authentication graphs provide intuitive insights into network user behavior. We believe that understanding these differences in even greater detail will lead to improved user behavior profiling and the elusive detection of authentication credential misuse.**

I. Introduction

User authentication is a fundamental aspect of modern computer use. This authentication can take the form of a simple username and secret password or involve more complex means of identity involving varying factors (e.g. biometrics, etc). However, these mechanisms nearly always become a unified authentication token of some form within a computer's operating system or application. This situation is particularly true within centralized authentication schemes where authentication tokens are cached and reused to access a variety of computers and services across the network. Most modern enterprise networks rely extensively upon centralized authentication systems with strong support from modern operating systems and applications. Kerberos is the most widely deployed example of a centralized authentication system. From a malicious insider's prospective, using either his or her own authentication credentials inappropriately or stealing others' is a necessary aspect of many malicious acts.

Existing work in profiling users for cyber security or other needs has primarily focused on host-based data sources and direct user actions [1], [2]. These studies and approaches do not consider actions of a user across a large set of computers. In addition, while there is research relating to graph analysis for network anomaly detection [3] and social networks [4], we believe our approach to using authentication graphs for analysis is novel. We see graphs providing an intuitive and extensive foundation for authentication activity analysis. The focus of the work presented in this paper is on improving the integrated security of all computers within an organization's network. We assume that enterprise defense relies on an overlapping set of approaches that includes the network interior. Given this assumption, the work presented here is about analyzing the significant authentication activity across a large population of users and computers that are part of a unified central authentication system.

We begin by describing the authentication system Kerberos [5], how it is implemented within an organization and how we create user authentication graphs from the system's authentication events. Next, using data from Los Alamos National Laboratory's (LANL's) centralized Windows-based Kerberos system (approximately 10,000 users), we present some initial analysis of user authentication graphs and their resulting security implications. More specifically, we provide an analysis of how privileged and non-privileged users differ, the usefulness of authentication graphs as an intuitive communications tool, and the use of authentication graph attributes as a predictor of user behavior. We consider this early research around user authentication graphs and will conclude with promising future work.

A. Centralized Authentication and Kerberos

Kerberos is the most widely deployed centralized authentication system, thanks to Microsoft's adoption of the system into its Windows operating environment (rebranded as Microsoft Active Directory authentication). Kerberos is a practical and scalable implementation based on the Needham and Schroeder symmetric key encryption network authentication protocol, which was proposed as one of the first examples of centralized authentication models [6].

The Kerberos protocol relies on a centralized server, referred to as a Key Distribution Center (KDC or Active Directory server in Windows), to be the centralized repository of trust (user authentication). Networked computers make authentication requests to the KDC, which provides trust delegation tokens (or *tickets* in Kerberos terminology). Computers, usually on the behalf of a user, make requests to the KDC for authentication tickets that are cached and reused automatically for the convenience of the user. The initial ticket, called a ticket granting ticket (TGT), becomes the locally cached authentication credential used to request future authentication tickets. The operating system or Kerberos-aware applications can use the user's cached TGT to make new requests to the KDC to get new authentication credentials that allow access to other computers or applications.

DOI 10.1109/SPW.2013.38

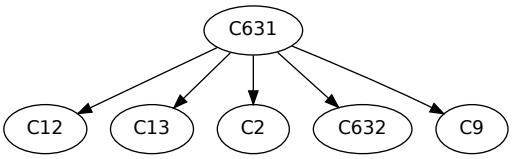

Figure 2. The network authentication graph from of a typical user *without* administrative access at LANL over 4 months in 2011. This user accessed 6 computers (nodes) total with a main computer (likely the user's desktop) connecting to 5 other computers. Note this creates a graph of diameter 1, a maximum out degree of 5, and a maximum in degree of 1. IP addresses have been mapped to avoid unnecessary information disclosure.

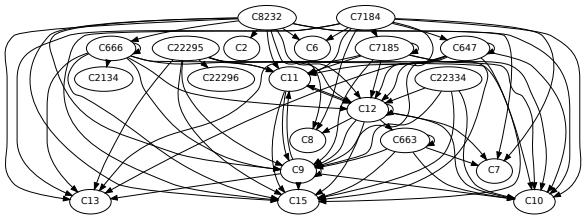

Figure 3. The network authentication graph from of a typical user *with* administrative access at LANL over 4 months in 2011. This user accessed 22 computers (nodes) total with 76 unique directional edges. Note this graph has a maximum diameter of 4, a maximum out degree of 11, and a maximum in degree of 11. It represents a significantly more complex authentication graph compared with nearly all non-administrative users.

These secondarily requested tickets are called ticket granting service (TGS) tokens. Both TGT and TGS tokens have finite lifetimes usually measured in hours or days (set as an enterprise policy within the KDC).

A standard set of Kerberos transactions for a user may look something like this: The user logs into a desktop computer with his usual username and challenge method (password, smartcard, etc), which causes the computer to request a TGT from the KDC. The TGT is successfully decrypted with the successful user challenge and cached in the computer's memory. The user then attempts and succeeds in mounting a network-base filesystem (share) from a server. This causes the desktop computer to use the user's TGT to request a TGS from the KDC for the fileserver. The KDC provides the TGS and the desktop then presents the specific TGS to the fileserver, which uses it to validate the user's authentication assertion to the filesystem. A TGS is requested for each specific application or computer on the user's behalf. This automated, repetitive process can easily get complicated. However, the key consideration for this research is that the KDC has full awareness of the user requesting a TGT from a specific computer and all subsequent TGS requests that include the source of the request and the intended destination (usually represented as network IP addresses). These pieces of information allow us to build an authentication graph of a user's activity based on data collected by the KDC. Two user authentication graphs using the TGT and TGS events recorded on a KDC for specific users can be seen in Figures 2 and 3.

II. Analysis of An Enterprise Authentication Environment

As the basis of exploration for an enterprise-sized centralized authentication system, we analyzed comprehensive centralized authentication activity over 4 months in 2011 from the Microsoft Windows Active Directory (KDC) authentication system at LANL. The total data set of more than 72 million successful TGT and TGS request records summarizes the activity of 9339 authentication user accounts using 22,368 networked computers all within a unified, single Kerberos trust domain.

Using the TGT and TGS events, agnostic of the user involved, creates a significant graph representing the aggregate authentication activity within the entire network. Over the 4 months, 204,838 unique, directed authentication edges spanned the network. The diameter of the graph was 21, the largest out degree from a computer was 703 (a configuration validation scanner), and the largest in degree was 14,709 (one of the Active Directory servers). Due to the substantial size and complexity, this unified authentication graph provides only basic analytical value: insights into potential credential mixing risks within the network and appreciation for the overall connectedness of the central authentication environment.

More interesting is the examination of specific user authentication graphs and the comparison of attributes between users. We have considered three categories of users to compare: users without any privileged or administrator access within the network, users with privileged or administrator access to one or more computers or systems within the network, and those users with institutional-level administrator access (the authentication master keys). Basic statistics over some key graph attributes are shown in Figure 1. Note that there are distinct differences within the three user categories showing administrative users having the much larger and more complex authentication graphs. These differences are also demonstrated visually in the typical non-administrator user authentication graph shown in Figure 2 and the administrator graph shown in Figure 3.

Through our analysis we find that host (node) count, graph diameter, and maximum in degree provide the most significant differentiators between user classes. Empirical probability densities for these three attributes can been seen in Figures 4, 5, and 6. Administrators have more complex graphs than typical users and the institutional administrators are even more complex.

The analysis of why administrators have more complex and extensive graphs is ongoing but some likely reasons can be hypothesized. For example, because administrators often manage a large number of computers, they are likely to log into many or all of those computers as a function

Attribute	Measure	All Users	Non-Administrators	Administrators	Inst. Administrators
Node count	Median	18	18	49	54
	Mean	21.82	19.88	67.22	84.50
	Std. Dev.	19.20	11.70	62.98	74.00
Edge count	Median	31	30	114	183
	Mean	53.01	45.14	247.84	321.15
	Std. Dev.	96.86	62.91	323.11	346.55
Diameter	Median	1	1	2	2
	Mean	1.27	1.24	2.08	2.25
	Std. Dev.	0.39	0.32	0.98	0.98
Max In Degree	Median	3	3	12	17
	Mean	5.66	4.86	25.41	27.65
	Std. Dev.	9.18	5.82	30.50	25.70
Max Out Degree	Median	13	13	19	27
	Mean	13.49	13.04	24.62	37.50
	Std. Dev.	6.75	4.73	22.19	54.56

Figure 1. A table comparing median, mean, and standard deviation for a variety of relevant graph attributes across each of the three categories of users (non-administrators, administrators, and institutional administrators) plus the total population of users. This data presents 9339 total users, 8957 non-administrators, 362 administrators, and 20 institutional administrators over a 4 month data set in 2011 from 72,697,000 total user authentication events logged.

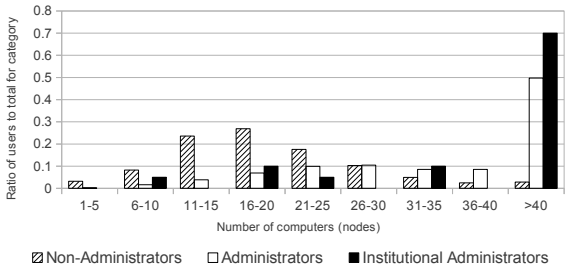

Figure 4. Empirical probability density for the number of unique hosts (nodes) each user authenticates to over a 4 month period in 2011 across the 3 user categories. Populations are 8957 for non-administrators, 362 for administrators, and 20 for institutional administrators. The majority of non-administrative users authenticate on few hosts compared to administrators.

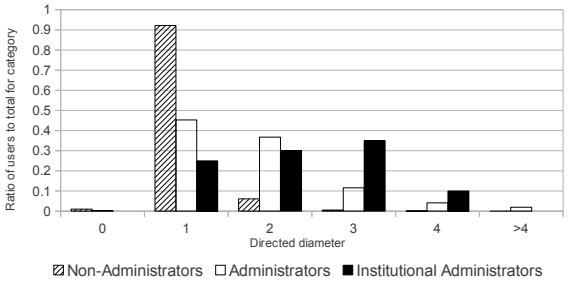

Figure 5. Empirical probability density for the maximum directed diameter of each user's authentication graph over a 4 month period in 2011 across the 3 user categories. Population sizes are the same as in Figure 4. Administrators tend towards longer diameters, indicative of more complex and chained authentication activity.

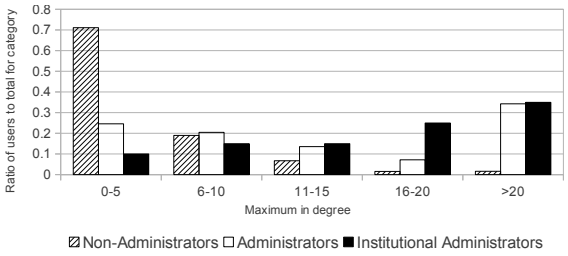

Figure 6. Emperical probability density for the maximum in degree of any node within each user's authentication graph from a 4 month period in 2011 across the 3 user categories. Population sizes are the same as in Figure 4. Administrators tending towards visiting central servers from multiple other hosts compared with non-administrative users.

of their job. They also rely on various network computers like central patch servers or application install servers that increase their authentication graph diameter and maximum in degree.

Of particularly interest are the few outliers with large and complex graphs who are *not* administrators. Only two users have diameter greater than 4. For maximum in degree greater than 20, 145 users need to be considered for additional scrutiny. Perhaps these users can be defined as power users or administrators that were missed in our classification method. While unlikely, maybe inappropriate behavior is being exhibited by some of these users (or by someone else misusing their tickets/credentials).

III. Risks and Results

Since having administrator access on a given host implies complete access to the system, it must be assumed that any other centralized credentials on the system may also be available to the administrative user. From an exploitation viewpoint, these authentication graphs become very enlightening. Should administrative access be available to a malicious actor on a given host, *all* users' authentication credentials cached on the host are available to the actor. Thus for hosts where multiple users authenticate (servers), should they become compromised, all associated users' graphs become a single, merged graph where all the combined nodes are now easily exploited. For credentials of administrators this vulnerability is of particular concern. Of course, the more locations the administrator authenticates, the greater the exposure.

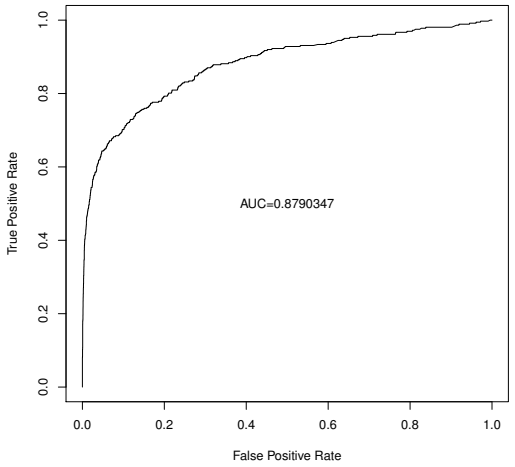

Figure 7. ROC curve showing the true positive and false positive trade-offs for determining whether a user is an administrator or not using a logistic regression model with the user's authentication graph characteristics and labels of administrators and non-administrators from the 4 month data set. True positives are successful predictions as administrators and false positives are inaccurate predictions as administrators. For example, our model shows that if we allow mislabeling 10% of users as administrators, we are able to correctly label approximately 80% of administrators.

The results of user authentication graphs involving administrators derived from this research has already resulted in operational changes in trust relationships for servers that have aggregated administrator credentials; servers that were previously not recognized as significant to security. In addition, the visualization of the institutional administrators' authentication graphs has increased awareness significantly through the demonstrated breadth of where high-value credentials are being exposed across the network. Based on the analysis presented in this paper, we are already seeing a change in behavior within this small group of central administrators, who are now more cognizant of credentials and the risks they pose as they are used through the network.

We have also considered the use of these authentication graph characteristics as a potential predictor of administrator or administrator-like behavior. To this end, we have developed a logistic regression model using the user authentication graph characteristics and the known label as an administrator or non-administrator. We believe the model, while still immature, has the potential to make useful prediction and determine inappropriate administrator-like behavior within the enterprise network. The model's resulting receiver operating characteristic (ROC) curve is shown in Figure 7. A similar model and results exists for institutional administrators (area under the curve$= 0.89$).

IV. Conclusions and Future Work

Looking forward, we see a significant continued opportunity for using user authentication graphs for analysis, user profiling, and visual representation. First, we see the need to move to time series analysis on the authentication graphs. While our static analysis has proven valuable, we see increased opportunity, fidelity, and practicality within various time series approaches. One particularly interesting analysis is to measure individual user graph variability over time, allowing us to use authentication graphs to detect malicious insiders. Next, we believe there would be significant value in understanding what outlier users are actually doing, as previously discussed. Understanding additional categories or subcategories beyond the three considered in this paper may be very useful. Finally, we see opportunities to mix the authentication graph characteristics with other user behavior measures to increase predictive quantification measures to operationally useful levels. For example, work has begun to integrate web browsing behaviors with the user authentication graph attributes to potentially predict and prevent computer and network compromise events.

This paper has demonstrated a useful and interesting way to examine and analyze large-scale authentication activity within a centralized authentication system. It has shown initial value in differentiating administrative and non-administrative users. It has also provided value as a tool in representing how wide-spread authentication activity can increase the risk of compromise to important centralized accounts. We see this work as only the beginning of a valuable research area.

References

[1] J. Iglesias, P. Angelov, A. Ledezma, and A. Sanchis, "Creating evolving user behavior profiles automatically," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 24, no. 5, pp. 854–867, 2012.

[2] D. Pepyne, J. Hu, and W. Gong, "User profiling for computer security," in *American Control Conference, 2004. Proceedings of the 2004*, vol. 2, pp. 982–987, IEEE, 2004.

[3] P. Ammann, D. Wijesekera, and S. Kaushik, "Scalable, graph-based network vulnerability analysis," in *Proceedings of the 9th ACM conference on Computer and communications security*, pp. 217–224, ACM, 2002.

[4] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, IMC '07, (New York, NY, USA), pp. 29–42, ACM, 2007.

[5] B. Neuman and T. Ts'o, "Kerberos: An authentication service for computer networks," *Communications Magazine, IEEE*, vol. 32, no. 9, pp. 33–38, 1994.

[6] R. Needham and M. Schroeder, "Using encryption for authentication in large networks of computers," *Communications of the ACM*, vol. 21, no. 12, pp. 993–999, 1978.