Kerberos Security With Clocks Adrift

Don Davis
Daniel E. Geer, Sc.D.

Kerberos Security
With Clocks Adrift

Don Davis* Daniel E. Geer, Sc.D.†

May 1, 1995

I used to be Snow White, but I drifted.

– Mae West

Abstract

We show that the Kerberos Authentication System can relax its requirement for synchronized clocks, with only a minor change which is consistent with the current protocol. Synchronization has been an important limitation of Kerberos; it imposes political costs and technical ones. Further, Kerberos' reliance on synchronization obstructs the secure initialization of clocks at bootstrap. Perhaps most important, this synchronization requirement limits Kerberos' utility in contexts where connectivity is often intermittent. Such environments are becoming more important as mobile computing becomes more common. Mobile hosts are particularly refractory to security measures, but our proposal gracefully extends Kerberos even to mobile users, making it easier to secure the rest of a network that includes mobile hosts. An advantage of our proposal is that we would not change the Kerberos protocol *per se*; a special type of preauthentication exchange can convey just enough replay protection to authenticate the initial ticket and its timestamp to an unsynchronized client, without adding process-state to the system's servers.

1 Introduction

The Kerberos Authentication System [19] provides password security for large networks. Unlike its principal competitors, KryptoKnight [12] and SESAME, [16] Kerberos requires that all of a network's system clocks must be synchronized. At first glance, this does not seem to be a great burden, at least for UNIX networks, but as Kerberos' influence has grown, synchronization has become a substantial impediment to Kerberos' adoption as a uniform networking standard.

Why has clock-synchronization become more difficult? We find that for three areas of explosive growth in the networking industry, there are good reasons for rejecting clock-synchronization. First, in-house wide-area networks have only recently become common. Corporate wide-area networks usually arise by agglomeration, so interdepartmental rivalries often obstruct centralized host management, including time-synchronization. Such practical political strains were less important in the more monolithic academic and engineering networks that adopted Kerberos early. It's clear, though, that monolithic networks are now passé, so Kerberos will have to accomodate such ordinary social tensions if its success is to continue. Second, online-access providers are bringing a massive surge of decentralized participants into the network industry. Obviously, it's commercially and technically infeasible to force synchronization on in-home network customers. Similarly, electronic commerce is bringing together buyers and sellers who share neither administrative nor organizational links; this openness guarantees that asynchronous clocks will remain the norm.

Third, the rise of mobile computing is bringing the problem of intermittent connectivity into renewed importance. Here, the problem is mainly technical: a time-synch protocol would burden both the laptop processor and its connection-initiation bandwidth. Here, a social obstacle to synchronization is that the intermittent user may alternate among several organizations' networks; synchronization would force such users' clocks to flutter. Mobile users are more vulnerable to security breaches than sequestered networks are, but laptops' intermittency and mobility obstruct the most effective approaches to open systems security:

- Intermittency obstructs time-synchronized cryptographic protocols.
- At the same time, challenge-response protocols

*Affiliations: Independent Consultant, 1318 Comm. Ave #16 Allston, MA 02134; *don@mit.edu*

†OpenVision Technologies, 1 Main St. Cambridge, MA 02142; *geer@cam.ov.com*

entail extra messages (see below), and these would impose unacceptable long-haul network delays on mobile users.

- Absent cryptography, the only alternative is firewalls and other protocol filters, but to a firewall, mobile users look like intruders.

This dilemma makes a synchronization waiver for Kerberos even more valuable. Once mobile users can authenticate themselves cryptographically, it becomes possible for a firewall or filter to recognize them, so that the home network can enjoy both styles of protection, instead of being denied both.

2 Why Synchronize?

In this section, we explain synchronization's purpose and alternatives that serve the same purpose, so as to review the history of synchronization's role in Kerberos' development.

Why does Kerberos need time synchronization in the first place? Synchronized clocks enable Kerberized applications to reject replay attacks. In a replay attack, an attacker eavesdrops on users as they present their credentials to servers; later, he resends the credentials to impersonate the users. A Kerberos client blocks replay by embedding an encrypted timestamp in each credential; the application server rejects credentials bearing out-of-date timestamps. Timestamps from users with slow clocks are indistinguishable from replays, so tolerating slow clocks gives attackers more time in which to work. Synchronization sharply limits this "replay window."

The alternatives to timestamping are all variations on "challenge and response." [7] In a challenge-response protocol, the credential recipient prevents replay by challenging each sender to encrypt and return a fresh random number, so as to demonstrate timeliness. The sender proves his identity by using his private key, or his session key, to encrypt the random number. Challenge-response protocols avoid the complication of synchronizing, but they always use at least one more message than a timestamp protocol, to accomplish the same security goal. Thus, it might seem that Kerberos' designers chose to optimize performance with timestamps and synchronization. As it happens, though, this speed/complexity tradeoff was *not* the reason Kerberos' designers chose a synchronizing protocol.

The 1978 Needham-Schroeder protocol, [13] from which Kerberos descends, used challenge and response to protect authentication credentials from replay. Three years later, Denning and Sacco [6, 3] pointed out that the N-S protocol was particularly vulnerable to compromised session-keys, because its key-distribution tickets made no provision for expiration of keys. They recommended that the tickets be timestamped, so that the session-keys would expire and be renewed regularly. They also recommended replacing challenge-response with timestamps in N-S' session-authentication handshake, and they pointed out that minute-resolution clock-synchronization would suffice to enforce key expirations. Here, synchronization helps to ensure that every connection gets a new session-key. This precaution makes it less profitable to steal session-keys or to attempt their cryptanalysis. Unfortunately, Denning and Sacco did not discuss the importance and difficulty of securing the time-synchronization process itself.

In the mid-80's, MIT's Project Athena incorporated Denning and Sacco's recommendations into their implementation of the Needham-Schroeder protocol, and added other protocols and security features, too. [19] With timestamping in place, N-S became Kerberos' flagship protocol, which we at Athena christened the "Authentication Service." This protocol handles all of Kerberos' password-mediated authentication, principally initial logins and password changes. Kerberos' other protocols enable a logged-in user to authenticate to additional services without entering a password anew, and without retaining the password on the local machine.

These newer protocols uniformly use encrypted timestamps to block replay, following Denning and Sacco's recommendation. However, Kerberos was designed to accomodate clock-skews of up to five minutes between clients and servers (though modern time services can synchronize much better than this). Thus, a replayed authentication-message will not be rejected as out-of-date, if it's less than five minutes old (a generous allowance, though not an unreasonable one). To close this security hole, Kerberos introduced a "replay cache," in which an application server stores each encrypted timestamp it receives for five minutes, the duration of the replay window. Each server should check every new timestamp it receives against its cache, so as to block replays of "fresh" timestamps.

In 1990, 12 years after Needham and Schroeder's paper, and five years after Kerberos' introduction, Bellovin and Merritt of AT&T Bell Labs wrote an important and insightful critique of Kerberos' version 4, which was influential in the design of the current version 5. [1] Along with other problems, Bellovin and Merritt pointed out that Kerberos security depends on *secure* clock-synchronization, and that V4 Kerberos was not itself sufficient to secure a clock-synchronization service. The clearest demonstration of this insufficiency is to consider a computer that is restarting automatically from a power failure, so that its system clock is certainly unreliable. In this situation, the computer cannot be sure of any message's freshness; indeed, if an attacker replays all of a previous day's network traffic, he can mislead the computer into using an old, compromised session-key as if it were fresh, and Kerberos' guarantees evaporate. As Bellovin and Merritt noted, the only way to defeat such an attack is with a challenge-response protocol, which Kerberos currently lacks, and which no current time-service supports.

It turns out that this situation is not merely illustrative, but is actually the crux of the problem. Only when a Kerberos principal first comes onto the net, does he need to use a challenge-response handshake to prevent credentials-replay. However, application clients and servers enter the network differently, so they must handle synchronization differently, too. Application servers need to use a challenge-response handshake only at bootstrap, to get time-service tickets. Thereafter, a server can trust its system clock, whenever it needs to renew its time-service tickets or other tickets it uses. For application clients, challenge-response is necessary whenever the user logs on to a physically-insecure workstation. Once the challenge-response handshake has assured the client of his initial tickets' freshness, the client does not need to synchronize his clock with the rest of the network. To be able to detect replay, the client only needs to know the difference, or skew, between his clock and the standard clock. [20] Thus, by adding a challenge-response handshake to only the Authentication Service protocol, we can break the circularity of Kerberos' dependence on a secure time-service.

3 Current Time Services

NTP is a cryptographically-hardened time service protocol. [10, 11] It enables a wide-area network to synchronize its software clocks with a few highly-accurate physical clocks. NTP's security has been extensively analyzed by Matt Bishop. [2] Each secure clock update depends on an uninterrupted chain of authentications, server-to-server, between the client and a remote physical clock. To mediate these authentications, NTP requires each host to maintain a shared key in a disk file, but makes no provision to distribute or refresh these keys. Kerberos can manage NTP's keys, but only under the assumption that the clocks are already synchronized. NTP makes no claim to solve this bootstrap problem; it assumes that secure key-management is available as reliable infrastructure, just as Kerberos assumes that time-synch is secure.

The Open Software Foundation's Distributed Computing Environment (OSF DCE) includes a secure Distributed Time Service, [15] whose security is mediated by DCE's Kerberos-based Security Service. For bootstrap, the DCE time service relies on the host's hardware clock chip to be physically secure, battery-powered, and accurate enough to fulfill Kerberos' secure synchronization needs. DCE explicitly accepts, just as Kerberos always has, that the clocks must be initialized "out-of-band," i.e., by wristwatch. [18] DCE's DTS is designed to interoperate with with NTP, but this interoperation does not address our bootstrap problem. Finally, neither NTP nor DCE's DTS makes any provision for physically-insecure hosts, which cannot hold long-lived keys on disk, and which therefore cannot participate in either protocol. Our proposal will work well with both of these services, without substantial change to their protocols or software.

4 Proposed Solution

In this section, we describe a "pseudo-preauthentication" protocol for Kerberos, that enables users to get tickets without having synchronized their clocks. We call this "pseudo-preauthentication," because we're abusing a flexible preauthentication extension, specified in Kerberos version 5. [14] Our protocol adheres to the specification, without obstructing true preauthentication. We also describe a mechanism that enables users to present accurate timestamps to Kerberos and to secure applications, without keeping their system clock synchronized. Unlike Bellovin and Merritt's suggested solution for Kerberos' synchronization problems, our proposals add no process-state to the Kerberos server or to the application servers.

True preauthentication proves the user's identity in his initial ticket request, so as to prevent attackers from requesting credentials in the user's name and attempting their decryption with a dictionary of commonly-chosen passwords. Bellovin and Merritt's paper included the first published analysis of Kerberos' vulnerability to this type of attack, and preauthentication is one of the solutions they suggested. There are many possible ways for a Kerberos client to preauthenticate his initial ticket request. [1] In the simplest and least secure way, the login client prompts the user for his password be-

[1]Most Kerberos suppliers have reinforced their Kerberos servers with password-quality controls, which arguably can prevent guessing attacks more definitively than preauthentication can do.

fore preparing the ticket request, and uses the password to encrypt a timestamp that authenticates the ticket request to the Kerberos server; the server refuses tickets to clients whose preauthentication fails. Proposals abound to overcome the flaws in this naïve scheme, employing both hardware and exotic software-only protocols, and the Kerberos version 5 specification made flexible provision for vendors to add any and all of these variations to the MIT implementation. [17, 14] To accomodate this variety, the specification document simply allows the client and server to include arbitrary, typed "preauthentication data" elements in their initial correspondence, and we shamelessly exploit this vagueness in the specification. The protocol allows unrelated types of preauthentication data elements to appear in the same message, so our use of the preauthentication option does not obstruct the simultaneous use of smartcards or some other other preauthentication scheme.

For clarity's sake, let's consider first how to initialize a clock securely, on a machine that does intend to synchronize. Suppose Bob is a system server who shares a key K_b with the Kerberos Authentication Server AS, and suppose he is willing to synchronize his clock. Every time he reboots, one of his tasks will be to request tickets for a secure time service S_t. To do this, Bob will send a nonce N_b in a challenge-response handshake:

$$\begin{aligned} B \rightarrow AS &: \quad B,\ S_t,\ N_b && (1)\\ AS \rightarrow B &: \quad T_{bt},\ \{S_t,\ L,\ K_{bt}\}^{K_b}, \\ & \quad \{N_b\}^{K_{bt}} && (2) \end{aligned}$$

The AS returns to Bob a new session-key K_{bt}, the key's times of creation and expiration $L = (L_{create},\ L_{expire})$, a ticket $T_{bt} = \{S_t,\ L,\ K_{bt}\}^{K_t}$, and the nonce N_b, newly encrypted. Except for the nonce components, Bob's exchange is identical to a usual Kerberos initial ticket request. Essentially, we've just conflated a challenge-response handshake into the standard protocol, formatted as preauthentication data. Note, though, that this handshake does not serve the usual preauthentication purpose of identifying Bob to the Kerberos server AS; instead it proves to Bob that the key K_{bt} and ticket T_{bt} are fresh.

Bob's nonce N_b is a random number, which he can generate from disk-drive randomness [5] or from some other noise source. It is important that Bob's choice for N_b must be immune to external influence; if an attacker can cause Bob to re-issue an old challenge N_{old}, then she can replay correspondingly old credentials T_{old}, $\{S_t,\ L_{old},\ K_{old}\}^{K_b}$, whose session key K_{old} she knows by prior theft. On receiving his time-service tickets from AS, Bob decrypts them with his password K_b, and uses the session key K_{bt} to decrypt the response to his timeliness challenge N_b. When Bob finds that the response is indeed his nonce N_b, encrypted with K_{bt}, he concludes that AS prepared the tickets after receiving N_b. As long as he has never used N_b to request tickets before, this means that the tickets are fresh. At this point, Bob can trust the tickets to afford secure clock-updates, or he can just use L_{create} to reset his clock immediately.

As in the usual Kerberos protocol, AS learns nothing from this exchange about whether it really was Bob who requested tickets, unless other preauthentication data authenticate him. Note though that when true preauthentication is available, it may make our challenge-response unnecessary. Many preauthentication mechanisms, such as smart-card protocols, were originally designed as mutual-authentication schemes in their own right, and do authenticate the server to the client. As long as the preauthentication protocol also protects the initial Kerberos credentials from replay, the client can trust the creation-time to represent Kerberos' current clock-time, without having to use our pseudo-preauthentication handshake.

Now, when Bob uses his new time-service tickets, he sends the usual authenticator, or encrypted timestamp, but it will probably be invalid:

$$B \rightarrow S_t \quad : \quad T_{bt},\ \{B,\ wrongtime\}^{K_{bt}} \qquad (3)$$

Note that in practice, Bob will probably put his tickets' recent creation-time into the authenticator, so *wrongtime* won't actually be far off. However, the timestamp doesn't have to be correct, anyway, because the time service doesn't care about his identity, and it doesn't worry about replayed requests. The time server's response returns the correct time *t.o.d.*, together with the request's timestamp:

$$S_t \rightarrow B \quad : \quad \{\mathit{wrongtime}, \text{“time}: t.o.d.''\}^{\mathrm{K_{bt}}} \qquad (4)$$

The first part of the server's response assures Bob that the time-report is fresh, because it echoes the timestamp he sent.

Suppose now a user Alice wishes to communicate securely with Bob, but suppose that like most users, she prefers *not* to synchronize her clock. In this case, Alice won't request time-service tickets, but she still needs to keep track of the Kerberos server's clock-value, so that she can prepare acceptable credentials, detect replays herself, and anticipate her tickets' expiration. Her ticket's lifetime data L tell her the current value of Kerberos' clock, because L includes the ticket's creation-time L_{create}. Alice can record the skew $\Delta_a = L_{create} - time_a$ between her clock and Kerberos', so as to keep track of Kerberos' clock's value. This fixed skew will enable her to prepare acceptable credentials, etc. as usual. [2]

Alice begins her login-session by asking AS for a ticket-granting ticket (TGT), which she'll then use to request tickets for Bob's service:

$$\begin{aligned} A \rightarrow AS \quad &: \quad A,\ TGS,\ N_a && (5)\\ AS \rightarrow A \quad &: \quad T_{a,tgs},\ \{TGS,\ L,\ K_{a,tgs}\}^{K_a},\\ & \quad\ \ \{N_a\}^{K_{a,tgs}} && (6) \end{aligned}$$

This is the same challenge-response handshake that Bob used above, except for the names. On receipt, Alice concludes that her ticket and session-key are fresh, just as Bob did, and she uses the key's creation-time L_{create} to construct a normal TGS ticket-request:

$$\begin{aligned} A \rightarrow TGS \quad &: \quad B,\ T_{a,tgs},\ \{A,\ L_{create}\}^{K_{a,tgs}} && (7)\\ TGS \rightarrow A \quad &: \quad T_{ab},\ \{B,\ L',\ K_{ab}\}^{K_{a,tgs}} && (8) \end{aligned}$$

Now, to detect replayed TGS-replies, Alice can compare her new ticket's creation-time L' with $time_a + \Delta_a$, which will be a good approximation to $time_{AS}$.

Note that after her initial login with the challenge-response, Alice's other security interactions are perfectly standard, and the rest of the Kerberos protocol is unchanged. However, to support drifting-clock clients, the Kerberos application library would have to be changed to maintain transparently an implicit "session clock" at each end of a Kerberized connection. Each side's skew $\Delta_{local} = time_{AS} - time_{local}$ would be initialized at login or at bootstrap; thereafter, whenever the Kerberos library needs synchronized time, it would add the skew to the local clock. This would allow an application's client and server to use Kerberos for security, even though neither party has synchronized his clock with Kerberos. Similarly, when a client interacts with several Kerberos servers, he'll have to maintain a separate clock-skew for each one.

Because the Kerberos protocol is unchanged, the drifting-clock clients and synchronized clients would be indistinguishable in their network behavior. Drifting-clock Kerberos clients and servers would fully interoperate with a normal Kerberos installation. If a drifting-clock client requests initial tickets from a Kerberos server that doesn't support challenge and response, the server will reject the preauthentication data. Then, the client can either initialize its session clock on faith, or it can reject the server's tickets as inauthenticable.

[2] Stan Zanarotti, of Dimensional Insight, Inc., devised an unsecured version of this clock-skew trick at MIT, when he implemented MIT's Kerberos clients for the Apple Macintosh. The trick is particularly necessary for the Mac, whose clock is hard to keep synchronized for a variety of reasons. [20]

5 Conclusion

We have presented a solution to Kerberos' "cold-start" problem in clock synchronization, which provides for secure clock initialization where needed, and for "drifting clock" security where desired. We expect the proposal to gain acceptance rapidly in the broad community of Kerberos' vendors, implementors, and designers, because it requires only minor changes to the Kerberos client library and to the secure time protocols, and because it adds no extra network delays to users' login sequence. Indeed, for Kerberos implementations that already employ preauthentication to protect against dictionary attacks, our proposal requires little more than a shift in interpretation, to exploit the fact that with some preauthentication schemes, Kerberos tickets already can be trusted to deliver a secure clock-initialization.

We also expect our proposal, once it's implemented, to greatly improve Kerberos' attractiveness to a variety of commercial network customers and users. Our notion of relaxed, yet secure, synchronization will further lighten administrative burdens and enhance security in large networks. It actually reduces Kerberos' administrative overhead, since most client machines will be able to dispense with time daemons, and it adds neither overhead nor network-latency to secure applications.

Intermittency, more than anything else, is the core technical challenge of mobile computing, yet mobile, intermittently connected counterparties have a bigger stake in authenticity than do continuously connected, sequestered network environments. As such, we claim that providing a solution easing Kerberos' synchronized clocks constraint is uniquely valuable because it enables the efficiency and prompt, assured revocation of authority (that is the hallmark of Kerberos authentication) to be broadly applicable to environments that do not and will not have time synchronization services. More broadly, we suggest that as the demands of electronic commerce become better understood, the ability to bridge the boundaries of internally synchronized yet mutually unsynchronized organizations will be shown to have compelling value.

6 Acknowledgments

We thank Barry Jaspan, Marc Horowitz, Jonathan Kamens, John Kohl, Joe Pato, Ted Ts'o, and Stan Zanarotti, for their helpful comments and suggestions.

References

[1] S.M. Bellovin and M. Merritt, "Limitations of the Kerberos Authentication System," in *USENIX Conference Proceedings*, pp. 253–267 (Dallas, TX; Winter 1991). Also in *ACM Comp. Comm. Rev.*, **20**(5), pp. 119–132 (October 1990). [research.att.com:dist/internet_security/kerblimit.usenix.ps]

[2] M. Bishop, "A Security Analysis of the *NTP* Protocol," *Sixth Annual Computer Security Conference Proceedings*, pp. 20–29 (Dec. 1990; Tuscon, AZ), [louie.udel.edu:/pub/ntp/doc/security.ps.Z]

[3] Michael Burrows, Martín Abadi, and Roger Needham, "A Logic of Authentication," *Proc. R. Soc. Lond. A* bf 426(1989) pp. 233-271.

[4] D, Davis and R. Swick, "Workstation Services and Kerberos Authentication at Project Athena," *Technical Memorandum* **TM-424**, MIT Lab. for Comp. Sci. (February 1990).

[5] D. Davis, P.R. Fenstermacher, and R. Ihaka, "Cryptographic Randomness from Air Turbulence in Disk Drives," in *Advances in Cryptology – CRYPTO '94*, Ed. by Yvo G. Desmedt. Springer-Verlag Lecture Notes in Comp. Sci. **839**, pp. 114-120 (1994).

[6] D. Denning and G.M. Sacco, "Timestamps in Key Distribution Protocols," *CACM* **24**(8), pp. 533–536 (August 1981).

[7] Li Gong, "Variations on Message-Replay Detection", 1992. (citation incomplete for now).

[8] Li Gong, "A Security Risk of Depending on Synchronized Clocks", *ACM Op. Sys. Rev.*, **26**(1) pp. 49–53 (1992).

[9] S.P. Miller, B.C. Neuman, J.I. Schiller, and J.H. Saltzer, *Project Athena Technical Plan*, Sec. E.2.1: "Kerberos Authentication and Authorization System," (Cambridge, Mass.) M.I.T. Project Athena internal document, Dec. 21, 1987.

[10] D.L. Mills, *Network Time Protocol (Version 2) Specification and Implementation*, Internet Request For Comments 1119 (Sept. 1989).

[11] D.L. Mills, *Internet Time Synchronization: the Network Time Protocol*, Internet Request For Comments 1129 (Oct. 1989).

[12] R. Molva, G. Tsudik, E. Van Herreweghen, and S. Zatti, "KryptoKnight Authentication and Key Distribution System." [jerico.usc.edu:pub/gene/kryptoknight.ps.Z]

[13] R. M. Needham and M. D. Schroeder, "Using Encryption for Authentication in Large Networks of Computers," *CACM*, **21**(12), pp. 993-999 (December, 1978).

[14] C. Neuman and J. Kohl, *The Kerberos Network Authentication Service (V5)*, Internet RFC 1510, September 1993.

[15] Open Software Foundation, *OSF^{TM} DCE Version 1.0, DCE Administration Guide* Volume 1, Module 4: "DCE Distributed Time Service," Rev. 1.0, Update 1.0.1. (Cambridge, MA; July 1992).

[16] T.A. Parker, "A Secure European System for Applications in a Multi-Vendor Environment (The SESAME Project)," *Proc. 14th Am. Nat'l. Sec. Conf.* 1991.

[17] J. Pato, *Using Pre-Authentication to Avoid Password Guessing Attacks*, (Cambridge, Mass.) M.I.T. Project Athena (December 1992).

[18] J. Pato, personal communication.

[19] J.G. Steiner, C.Neuman, and J.I. Schiller, "Kerberos: An Authentication Service for Open Network Systems", *USENIX Winter Conference Proceedings*, February 1988. [athena-dist.mit.edu:pub/kerberos/doc/usenix.PS]

[20] S. Zanarotti, of Dimensional Dynamics, Inc. was the first to use this trick; personal communication.