公用事业自动化收费系统数据安全性的研究 [1]

龙子庄 李声沛

(中国科学院电子学研究所 北京 100080)

摘 要 该文简要介绍了公用事业自动化收费系统的基本构成，并就系统的数据安全性进行讨论，提出了一个“以随机数确定数据传输顺序，插入无效随机数增大信息冗余，用哈希函数校验，用一次性加密方法加密哈希值和随机数”的数据安全解决方案，有效防止系统数据遭受主动攻击.

关键词 自动化收费系统，数据安全，随机数，哈希算法，一次性加密

中图号 TN99

1 引 言

近年来，远程数据采集传输通信技术得到了长足的发展，形成了电力载波、无线、485 总线方式通信等应用技术领域。这些技术在解决低成本高效率的远程仪表数据采集传输和智能小区管理智能化等系统中得到了广泛应用。经过不断的发展，远程数据采集传输通信技术水平不断提高，已逐渐产品化。然而，远程数据采集传输通信系统特别是水、电、燃气等公用事业自动化收费系统的数据安全问题没能引起很好的重视。现有的系统基本上没有采取数据安全措施，明文传输，或是仅采用简单的循环冗余校验 (Cyclic Redundancy Check, 简称 CRC) 或变体的 CRC 校验。虽然这些系统的安全性能没有银行网络等要求高，但是如果这个问题解决不好，就会存在严重的安全隐患，可能造成严重的经济损失，并产生不良的社会影响。

一般来说，一个信息系统的安全度，体现为反泄密、反窜改和反破坏的能力。除信息存储和处理过程外，信息在传输过程中受到的安全威胁最大。安全威胁大体可分为被动攻击和主动攻击两类 [1]。被动攻击旨在破坏信息机密性，主动攻击旨在破坏信息真实性和完整性。就公用事业自动化收费系统而言，被动攻击就是非法入侵者截获并解密通信系统传输的数据，而主动攻击就是非法入侵者在获取数据的基础之上，设法将系统传输的数据进行篡改。本文通过开发具体的远程收费系统，对其数据安全性从防被动攻击和防主动攻击两个方面作了全面深入的研究，提出了一种适合于运算能力低的系统的综合数据安全方案，即“以随机数确定传输数据顺序、用单向哈希函数对数据进行运算、用一次性加密方法对哈希值和随机数进行加密”。该方案简单实用，高效可靠。

2 系统构成

公用事业自动化收费系统主要由监控中心、面向小区的集中器、面向用户的采集器和计量仪表 (如水表、电表、燃气表等) 组成，如图 1 所示。监控中心是由计算机、网络管理软件和数据库构成，主要负责定期读取集中器中计量表的数据和状态，根据需要通过集中器和采集器设置修改计量表的参数，并通过互联网 (Internet) 与有关银行的网络连接，传输用户数据，实现自动缴费。集中器定期读取采集器中计量表数据，同时兼作采集器与监控中心之间通信的桥梁。采集器负责实时采集记录计量表的脉冲数，并且检测用户计费表的状态。

3 系统的安全威胁

公用事业自动化收费系统使用的是公开的共享传输媒介，任何用户都可以无限制的将设备接到网络上。这就使得数据安全更容易受到威胁。考虑到网上传输的数据是用户的真实消费数

[1] 2001-05-08 收到，2001-07-19 定稿

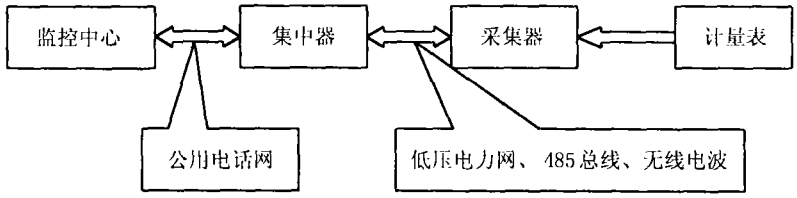

图 1　系统构成

据，非法查看这些数据不会直接带来不良影响，然而截获数据是篡改数据的基础。系统的安全威胁主要是来自非法入侵者对用户数据的非法篡改，使得监控中心读取的用户数据是错误的。安全威胁主要是发生在集中器向采集器采集数据的时候，系统遭入侵前后采集器到集中器的数据流向示意图如图 2 所示，有两种不同的表现形式：

(1) 当非法入侵者获得集中器发出的命令后，并截获采集器发出的数据，然后将数据进行篡改再发给集中器；

(2) 在接到集中器发出的采集命令后，将过去截获的数据发给集中器。

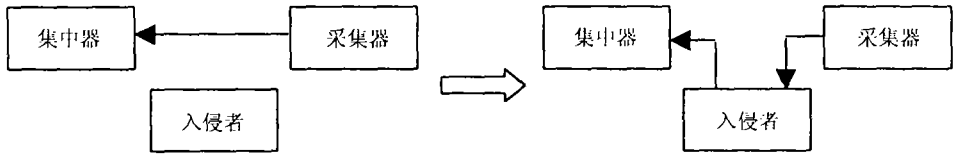

图 2　系统遭入侵前后采集器到集中器的数据流向示意图

4 数据安全方案

4.1 多种加密方法抵御非法入侵者对数据的篡改

在公用事业自动化收费系统中，集中器和采集器都是采用普通的单片机，它不可能进行复杂的加密解密运算。所以现在流行的加密方法无法应用于该类系统。在借鉴计算机网络的加密技术并在深入研究本系统特点的基础上，提出如下简单高效的综合解决方案。

针对安全威胁的第一种情况，我们只要防止主动攻击，即保持从采集器传到集中器的数据的完整性。我们可以这样设想，在需要传输的消息 (M) 后面加一段识别码 (I)，对于不同的消息 I 的值是不一样的。接收方将收到的消息按相同的公式进行计算，若得到的 I 值是相同的，则证明所收到的数据是完整的。相反则将数据丢弃并要求发送方重发。我们可以从两个角度考虑，一是让非法入侵者无法识别有效数据，二是即使其能识别有效数据也使其不可能准确得到识别码 (I)。只要这两种情况中任何一种成立，系统就是安全的。

本文提出的方案原理如图 3 所示。图中 x 是要传输的明文，发送方首先对 x 进行哈希运算，接着用随机数控制 x 进行顺序重排，最后用一次性加密方法对哈希值和随机数进行加密，加密后的哈希值与随机数和排序后的明文一起传输。接收方在接到数据后，先将随机数和哈希值进行解密，然后根据随机数将接收的数据恢复原序，得到明文 x'，对 x' 进行哈希运算，若得到的 $h(x')$ 和接收到的 $h(x)$ 相等，则证明接收到的数据是正确的，否则将数据丢弃并要求发送方重传。具体从如下 4 个方面进行阐述：

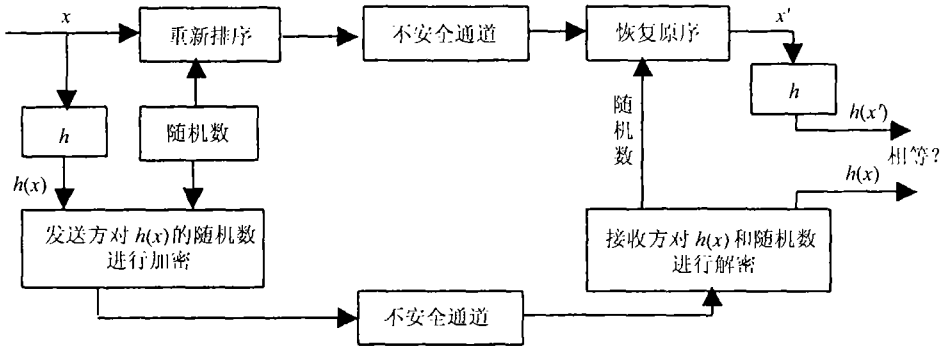

图 3 多种混合加密方法原理图

4.1.1 *用随机数确定传输数据的排列顺序* 每次数据传输的顺序都是随机的，同时因为数据不具有现实自然文的意义，非法攻击者要在短时间内找到相应的消息顺序是不可能的。

该方法的思想是在通信开始之前，将各种可能的排序预先设计好并编号 (设共有 m 种排列方法)，然后存于通讯双方设备的数据存储器 (E^2PROM) 中。通信进行时，发送方根据随机数查询对应的排序，确定传输数据的顺序，在传输数据的同时，将确定传输数据顺序的随机数进行加密随数据同时传输，在接收方接收到数据后，根据接到的随机数查取数据的顺序，然后恢复成正常的顺序，再根据需要对数据进行相应的处理。现在关键问题在于如何产生随机数。

一个简单而高效的方法是从实时环境中取得。在远程数据采集系统中，可由计费仪表中的脉冲累加器提供。因为，脉冲产生是随机的，自然而然，其累加器的脉冲总数也是随机的，因此脉冲累加器中的脉冲数可以作为随机数使用。设脉冲累加器的阈值为 T(即最大值)，脉冲累加器中的脉冲数为 n。由 m 和 T 的关系不同，存在 3 种情况：

(1) $m = T$， n 直接作为随机使用。

(2) $T > m$ 时，就有可能存在 $n > m$ 的情况，可将 n 对于 m 的模，即 n 为 $\mathrm{mod}(m)$ 得到最后的随机数。

(3) $T < m$，直接将 n 作为最后使用的随机数就不能充分地发挥系统的性能，记 $D = m - T$，计算 n 为 $\mathrm{mod}(2)$，若为 0，将 n 作为最后的随机数，若为 1，将 $D + n$ 作为最后的随机数，这样就可以有效地覆盖所有排列的取值。

对于长度为 m 字节的数据，不同的排列顺序共有 $m!$ 种。若 m 很大时，存储的编号顺序就会占用很大的空间。这时可以考虑只对整个序列中部分的序号进行重新编号。比如当 $m = 20$ 时，只对第 0，4，8，12，16 等字节进行序号重排。这时只有 5!=120 种。当然，参与重排的序号数量可以随安全性能要求的变化而变化。

4.1.2 *传递无意义的随机值，增加信息冗余度，增大数据解密的复杂程度* 在传输的数据中的一些固定位置上安插一些没有意义的值，这些值每次由系统随机产生，因此其每次变化会很大。由于传输的数据中有意义部分也不具有现实自然语言的意义，非法入侵者很难辨别数据中那些是有用的信息，那些是无用的信息，因而也无从对数据进行任何有意义的修改。

这些随机数也采用上面提到的产生随机数的方法。按在预定的位置填上脉冲累加器中脉冲数。一般情况下，每个采集器都会采集多个表的数据，因此也就存在多个随机数。若随机数不够，也可以对随机数进行取模或作 “异或” 运算取得。

4.1.3 *用哈希算法对传输的数据进行运算* 运用哈希函数算法加密保持数据完整性是一种常用的加密方法。它具有以下的特点：(1) 由输入得到输出是容易的，也就是算法是简单的；(2) 已知算法的输出，要求得到的输入是困难的；(3) 随机性，即输入相同时，输出一样，而输入的每一个比特的变化都对最后 I 值的影响很大。具有上述特性的算法是哈希函数。然而，以往的

哈希算法都是针对高速通信系统设计的，对低速系统来说算法过于复杂。本文针对低速系统，采用简化的哈希算法，以提高效率，图 4 是单向哈希函数示意图。

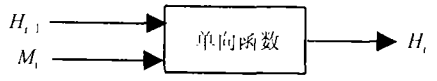

图 4　单向哈希函数

该哈希函数是建立在给定两个长为 n 的输入、输出一个长为 n 的值的单向函数之上 [2]。函数的一般输入为消息块和前一块消息的哈希值，其关系式可表述如下：

$$H_i = f(M_i, H_{i-1})$$

最后一块消息的哈希值作为整个消息的哈希值。这样无论输入的大小如何，单向哈希是形成固定长度的输出。

算法描述如下：

对于给定的串 X 将其分解成以下形式的串：

$X = X_0X_1 \cdots X_{n-1}$

$X_i(0 \le i \le n-1)$ 是长为 2 个字节 (16 位) 的子串，最后子串不足 16 位的，补至 16 位。

定义以下运算：

$f(x,y,z) = (x \wedge y) \vee (\overline{x} \wedge z)$ 其中 $\wedge$ 为 "与" 运算，$\vee$ 为 "或" 运算，$-$ 为反运算

$g(x,y,z) = (x \wedge y) \vee (x \wedge z) \vee (y \wedge z)$

$h(x,y,z) = x \oplus y \oplus z$，其中 $\oplus$ 为 "异或" 运算

$F(x,y,z)$ 表示 $x = (x + f(x,y,z)) << 1$,

$G(x,y,z)$ 表示 $y = (y + g(x,y,z)) << 3$,

$H(x,y,z)$ 表示 $z = (z + h(x,y,z)) << 5$, 其中 $<<$ 为循环左移

初始化如下，例如设 y=01234567, z=FEDCBA ，做如下循环计算：

For $I = 0$ to $n-1$ do;

$x = X_i$

$F(x,y,z), G(x,y,z), H(x,y,z)$

Next

最后得到的 y, z 值组合成一个 8byte(64bit) 的哈希值。哈希值经第 4 步加密后，与其它数据一并发送。

4.1.4 *用一次性的加密方法对随机数和哈希值进行加密*　使用一般的私钥或公钥进行数据加密，对于运算速度慢的系统来说是一个极不可取的。而一次性的加密技术已被发现多年，并被证明为可以轻易地设计出不可被破译的密码 [3]。它的主要思想是选择一个随机比特串作为密钥，然后把需要加密的明文变成比特串，最后逐位对两个比特串作 "异或" 运算。由于每一段明文都同样是密钥，即作为密钥的比特串和需加密的比特串对于最后形成的密文的作用是平等的。这种密文没有给破译者提供任何可能的信息，因而无法破译这样的密文。

虽然使用这种加密方法进行加密的密文很难被破译，但现在运用得却非常的少，主要原因有：(1) 这种密钥无法记忆，故收发双方都需随身携带密码本。任何一方造成密码本丢失都会造成全部泄密；(2) 可传送的数据总量受可用密钥的数量限制。然而，一次性加密的方法却非常适合远程数据通信系统的特点。理由有：(1) 需要加密的数据只包括哈希值和随机数，数据量不大。(2) 另一方面集中器和采集器是统一安装的，所以可以预先把作为加密的密钥的随机数固化到设备上。(3) 方法简单。整个加密的过程只进行 "异或" 及其组合运算，对于单片机控制的低速系统来说非常的合适。加密的示意图如图 5 所示。

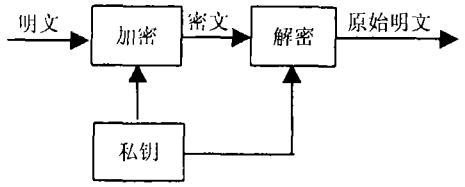

图 5 加密示意图

运用一次性的加密方法成功的关键在于对作为密钥的随机数列进行保密。因为系统的源程序只有少数人能够看见，因此对密钥的保密是可能的。

4.2 用时间戳防止非法入侵者利用过时的消息进行攻击

针对第二种情况，我们只需将要发送的消息加盖一个时间戳，也就是把日期和时间作为输入，经上述的算法运算后，得到的值作为时间戳。这就可以保证集中器每次收到的数据都是采集器的最新数据，而不是很长时间以前的数据。

5 结 论

本文讨论的公用事业自动化收费系统数据安全性研究，通过实际测试和使用，证明对系统数据的防主动与被动攻击都非常有效，主要体现在以下几个方面：(1) 单向的哈希函数使得非法入侵者篡改数据可以及时发现。(2) 按随机的方式打乱数据的序号，加大了非法入侵者作案的难度。(3) 在数据中加入无意义的随机值，增加信息的冗余度，加大非法入侵者篡改有意义数据的难度。(4) 使用一次性的加密方法，是非法入侵者对关键的哈希值和随机数无法进行解密。(5) 使用时间戳，使非法入侵者利用以前的数据作为当前的数据能被及时的发现。

本文讨论的公用事业自动化收费系统数据安全方案具有普遍性，可适应多种实际需求，对同类系统的开发和设计具有一定的实用参考价值和指导意义。

参 考 文 献

[1] 卢开澄，计算机密码学—计算机网络中的数据保密与安全，北京，清华大学出版社，1998 年，第 1 章.

[2] 冯晖，来凤琪，王绍银，章群，计算机密码学，北京，中国铁道出版社，1999 年，第 14 章.

[3] Andrew. S. Tanenbaum, 熊桂喜，王小虎译，计算机网络，北京，清华大学出版社，1998 年，451–452 页.

STUDY OF DATA SECURITY IN FEE DATA AUTOMATICALLY COLLECTING SYSTEM OF PUBLIC UTILITY

Long Zizhuang Li Shengpei

(*The Institute of Electronics, Chinese Academy of Science, Beijing* 100080, *China*)

Abstract The structure of fee data automatically collecting system in public utility is introduced. With the stuty of the security of the system, a synthetic security scheme is put forth, which is using random number to decide the sequences of the data to be transmitted, inserting random numbers with no use to improve information redundance, using monodirectional hash function to get the exclusive value of the transmitted data and adopting the one time encrypt to encipher the hash value and the random number. This method is simple and highly efficient.

Key words Automatically collecting fee data system, Data security, Random number, Hash function, One time encrypt

龙子庄： 男，1974 年生，硕士生，研究方向：信号与信息处理.

李声沛： 男，1939 年生，研究员，研究方向：智能控制与智能自动化.