ITNG 2018 Proceedings- ONLINE BOOK OF ABSTRACTS

A New Proposed Unconditionally Secure Blind Proxy Signature Scheme Based on RSA

*MohamedMohamedKouta,Arab Academy for science & Technology,mmkouta@gmail.com

*MagedElgendy,Arab Academy for science & Technology,melgendy562002@yahoo.com

*AbdElftahhegaz,Arab Academy for science & Technology,ahegazy@aast.edu

ABSTRACT

A proxy signature scheme, a variation of ordinary digital signature scheme, enables a proxy signer to sign messages on behalf of the original signer. Proxy signature schemes have been shown to be useful in many applications. For example, a manager can delegate his secretaries to sign documents while he is on vacation. We provide a new model for practical proxy signature using digital certificates. The proposed model is satisfied the security requirements like Identifiability, Unforgeability, Undeniability, Verifiability and prevention of misuse and also provide instantaneous revocation of delegation. If the original signer wants to revoke this delegation he sends a revocation request to the delegation authority and the delegation authority approves the revocation request and revoke the delegation by adding proxy public key to publish it in the revocation list

.

1. Introduction

Digital Signature technology ensures the authentication, integrity, privacy and non-repudiation and with the Internet has rapidly become more advanced and popular in recent year that has applied paperless and E-commerce applications so it must take digital signature into focus. Digital signature is applied in many application like electronic voting, electronic lottery, e-commerce and etc. Many special digital signatures such as blind signature, ring signature, threshold signature and proxy signature have developed the important branches of applied digital signature. If the signer/Manger was in vacation somewhere during holiday. However, the business should still continue for these days. Thus induce the requirement of a signer to delegate his/her signing Right to another user by applied proxy signature and the delegated user is called proxy signer.The first notion of proxy signature was introduced by Mambo, Usuda and Okamoto [1], and they divided the delegation signing into three types' full delegation, partial delegation and delegation by warrant.

A secure proxy signature model must satisfy the following requirements [2]:

1. Identifiability: Identity of a proxy signer and the original signer can be determined from the proxy signature.
2. Unforgeability: Only the designated proxy signer can create the valid proxy signature on behalf of the original signer.
3. Undeniability: Once the proxy signer generates the valid proxy signature on behalf of the original signer, cannot deny a signature creation against anyone.
4. Verifiability: From a proxy signature, the verifier is convinced of an original signer agreement on the signed message.

5. Prevention of misuse: it should be confident the delegation can't be used in another purposes.

Followed by first notion, number of models and improvements have been proposed; however, most of them do not fully meet all the security requirements of a proxy signature model [3-10]. Most proposed proxy signature models are based on discrete logarithm problems [3], [4], and [5]. Some proxy signature schemes are constructed from pairings [6,7]. None of the above mentioned proxy signature schemes have the instantaneous revocation capabilities. None of the proposed model do not use digital certificate and use private key generator (PKG) [10]. The PKG has the following of drawbacks:

1. If PKG is compromised, all messages protected over the entire lifetime of the public-private key pair used by that server are also compromised.
2. Because the Private Key Generator (PKG) generates private keys for users, it may decrypt and/or sign any message without authorization. This implies that IBE systems cannot be used for non-repudiation.

2. Demonstration of the Proposed Proxy Digital Signature Model

In our proposed model, four entities are involved, which are original signer (S), proxy Signer (P), Trusted Delegation Authority (T) and Verifier (V). The S delegate his signing right to P to sign the message on behalf of the original signer S. The V knows the public key of S and P can validate the proxy signature. The T is the third party, which issued the certificate to proxy signer that contains proxy public key and specify the kind of message are delegated. The delegation period and any time the S wants to terminate this delegation for P, the T can terminates instantaneously.

The proposed model consist of the following four stages:

1. Setup Stage: the proxy key pair is generated in this stage.
2. Signing Stage: The P is signing behalf of S.
3. Verification Stage: The V verifies the signature of P and S on the messages.
4. Revocation Stage: any time S wants to terminate his delegation for P.

A. Preliminaries

Let E={S,P,T,V} be set of entities involved in the proposed model where S is the original signer, P is the proxy signer, T is the trusted third party that is witnesses that the S delegates his signature to P, and V is the verifier that verify the signature of the proxy signature. Let X, Y € E, X is represented as follows.

$$X= \{pub_{x,} priv_x, n_x\},$$

where pub_x is the public key of entity X, $priv_x$ is the private key of X and n_x is the RSA constant.

$$Y= \{pub_Y, priv_Y, n_Y\},$$

where pub_Y is the public key of entity Y, $priv_Y$ is the private key of Y and n_y is the RSA constant. We assume the message M following holds:

$$(M_{AES}) \| ((((AES_{Session\ Key})_{RSA}{}^{pubx}) \bmod n_x)^{privx}) \bmod n_x =$$
$$(M_{AES}) \| ((((AES_{Session\ Key})_{RSA}{}^{privx}) \bmod n_x)^{pubx}) \bmod n_x =$$
$$(M_{AES}) \| (AES_{Session\ Key})_{RSA}$$

$$((M_{AES}) \Vert(((AES_{Session\ Key})_{RSA})_{RSA\text{-}1}))_{AES\text{-}1}$$

For two entities X, Y € E, Y ≠ X, a message M from X to Y is sent as follows:

$$((((H\ (M)^{privx})\ mod\ n_x)^{puby})mod\ n_y$$

Y decrypts and verifies the message as follows:

$$(((((((((H\ (M))^{privx})\ mod\ n_x)^{puby})mod\ n_y)^{privy})mod\ n_y)^{pubx})modn_x)= ((H\ (M)))$$

Thus we can represent the entities involved as follows:

1. $S= (pub_s , priv_s , n_s)$
2. $P= (pub_P , priv_P , n_P)$
3. $T= (pub_T , priv_T , n_T)$
4. $V= (pub_V , priv_V , n_V)$

B. The Proposed Model Stage

The proposed model consist of four stages, which are the setup stage, signing stage, verification stage and revocation stage.

1) Setup Stage

I. S performs the following:

1. S generates a large prime N_1Є $\{max\ [p_s, q_s] +1, \Phi\ (n_s)-1\}$ and computes N2 as the inverse of N_1 mod (n_s) for two large primes $p_s, q_s , n_s= p_s * q_s$
2. S creates the partial proxy key $N_3=N_1* priv_s$
3. S signs on (N_3, N_2) by his private key and creates F_1 as follows:
   $F1= ((N_3, N_2))^{privs}mod\ (n_s)$
4. S create an envelope

   $BSK= (F_1)^{pubp}mod\ (n_p)$,

Where BSK is the blind signature key and sends it to P.

II. P retrieves the partial proxy key as follows:

$F_1= (\ BSK\ )^{privP}mod(n_P) = ((\ F_1\ )^{pubp}\ mod(n_p))^{privP}mod\ (n_P)$,

$F_1= (N_3, N_2\ )^{privs}\ mod(n_s)$ & $(N_3, N_2) = (F_1)^{pubs}\ mod\ (n_s)$.

1. P also generates large prime N_5Є $\{max\ [p_P, q_P]+1, \Phi(n_P)\ -1\}$ and computes N_6 the inverse of N_5 mod (n_p) for two large prime numbers p_P, q_P, where $n_P= p_P * q_P$.
2. P creates the proxy key as follows:
   a) $N_7=N_5* priv_P$
   b) $N_7 * N_3 = N_5 * priv_P* N_1* priv_s =X_P$, where X_P is the private proxy key, (1) &
   c) $N_6 * N_2 = Y_P$ is the public proxy signature key and sends it to S. (2)
3. S sends delegation request to T
   The delegation request contains delegation period, message type and public proxy key.

2) Signing Stage

P creates proxy signature and sends it to T as follows:

$$(H\ (M))^{X}p\ )^{pubT}\ mod(n_T) = ((H\ (M))^{\ N1*privs\ *\ N5\ *\ privp\ })^{pubT}\ mod(n_T)$$

T decrypts and verifies the proxy signature by using public proxy signature key and checks M within context and authority of P as delegated from S. If P is dedicated with the policy of S, T will response via secure BCS channel to P by adding his signature as follows.

$((H(M))^{N1 * privs * N5 * privp})^{privT} \bmod(n_T)$.

If not T will not add his signature. P encrypts and sends via secure channel the response from T to V as follows:

$(((H(M))^{N1 * privs * N5 * privp})^{privT} \bmod(n_T))^{pubv} \bmod(n_v)$

3) Verification Stage

V removes the encryption of the proxy signature by his private key and verifies the signature of T, P and S as follows: V Verifies the signature of T

$$(((H(M)))^{N1 * privs * N5 * privp})^{privT} \bmod(n_T))^{pubT} \bmod(n_T)$$

$$= (H(M))^{N1 * privs * N5 * privp}$$

V verifies the proxy signature as follows:

$$((H(M))^{N1 * privs * N5 * privp})^{Yp} =$$

$$((H(M))^{N1 * privs * N5 * privp})^{N6 * N2} =$$

$$((H(M))^{privs})^{privp}$$

The final results is the original data which is signed by the original signer and the proxy signer. The verifier verifies the data signed by S & P as follows:

$$((H(M))^{privs})^{privp})^{pubp})^{pubs} = (H(M))$$

4) Revocation Stage

This section provides to S the ability to revoke the delegation at any time by sending to T via secure channel the revocation request that contains the proxy public key and public key of proxy signer. When P signs the message using proxy key and sends it via secure channel to T, the trusted third party, T, will not add his signature to proxy signature and this refers that this message is not trusted. T revokes the delegation by adding the proxy public key to the publishing revocation list.

3. The security of the proposed scheme

Security of our scheme presented in the next few lines

1. Forgery by the Original Signer:

   The proxy secret key is dependent on both the proxy information sent by the original signer as well as the secret key of the proxy signer. Therefore the original signer cannot generate the proxy secret key. He also cannot derive the proxy secret key from the proxy public key given by equation (2) as it is difficult to factorize the integer *N*. Thus the original signer is unable to sign like the proxy signer. Therefore forgery by original signer is computationally not possible.

2. Impersonating Attack:

   Let us assume that Bob is not designated as a proxy signer by the original signer Alice. Though Bob can generate a proxy key pair (*X1p; Y1p*) satisfying equations (1 and 2) and sign a message on behalf of Alice, the verifier on receiving the signatures, first confirms using a verification equation whether the signature is from a valid proxy signer or from a revoked

proxy signer. During this test the verification fails and the verifier considers him as a revoked signer. Thus Bob cannot become the proxy signer unless he is designated by the original signer Alice.

3. Framing Attack:

In this attack, a third party Charlie forges a proxy private key and then generates valid proxy signatures such that the verifier believes that these proxy signatures were signed by the proxy signer Bob on behalf of the original signer Alice. When such a proxy signature is presented, Alice cannot deny that she is the original signer of the proxy signer Bob. The result is that Alice and Bob will be framed to accomplish this attack, Charlie needs to forge Bob's proxy key pair (*Xp; Yp*). As forward-secure signatures are used by proxy signer it is computationally difficult to forge the proxy secret key. Knowing the proxy public key *Yp*Charlie cannot generate the proxy private key given by equation (2) as it is difficult to factorize the integer *N*. Thus our scheme withstands the above attacks. By this we can say that only the designated proxy signer can create a valid proxy signature on behalf of the original signer. In other words, the original signer and other third parties who are not designated as proxy signer cannot create a valid signature. Thus the second requirement, Strong unforgeability, of a secure proxy signature is satisfied.

4. The Advantages of the proposed scheme

The proposed scheme provides fives levels of key-hierarchy to achieve unconditionally secure hybrid scheme for sensitive and normal applications. The first is the pass phrase initialization vector used with the AES and SHA algorithms to encrypt the private keys. The second is the AES session key, which is used to encrypt the messages before sending them. The third is the RSA public key, which is used to encrypt the AES session keys. The fourth is the RSA public or private keys that used to encrypt or signing the parameters before sending. The fifth key-hierarchy is the keys used and generated from the standard cryptographic **Pseudo-Random bit Generator** ANSI.X9.17 for ultimate secure channels between the authorized parties in the bit commitment scheme. The scheme has been implemented on a commercial PC, allowing better portability, maintainability, and availability.

5. CONCLUSION

Proxy signature is an important delegation technique, it is widely used in auto-office system and it has potential commerce value in modern information world. The proposed protocol is a fully controlled delegation protocol with instantaneous revocation capabilities. The proposed scheme allows easy, simple, instantaneous revocation and satisfied the security requirements like identifiability,unforgeability, undeniability, verifiability .

References

1. Mambo M, Usuda K, Okamoto E. Proxy signature: delegation of the power to sign messages. IEICE Transactions on Fundamentals, 1996, 79-A (9): 1338_1353.

2. M.-S. Hwang, E. J.-L. Lu, and I.-C. Lin, A Practical (t, n) Threshold Proxy Signature Scheme Based on the RSA Cryptosystem. IEEE Trans. Knowledge and Data Engineering, vol. 15, no. 6, 1552-1560, 2003.

3. Kim S., Park S., & Won D. (1997) Proxy signatures. Revisited. In: ICICS'97. LNCS 1334.Springer-Verlag, 223–32.

4. Sun H-M, Lee N-Y, & Hwang T. (1999).Threshold proxy signatures.IEE Proc – Comput Digit Tech, 146(5), 259–63.
5. Sun H-M. (2000). Design of time-stamped proxy signatures with traceable receivers.IEE ProcComput.Digit. Tech.,147(6), 462–6.
6. Zhang F, Safavi-Naini R, & Lin C-Y. (2003). New proxy signature, proxy blind signature and proxy ring signature schemes from bilinear pairing. In: Information security andprivacy (ACISP'03), LNCS 2727. Springer-Verlag, 312–23.
7. Okamoto T, Inomata A, & Okamoto E. (2005). A proposal of short proxy signature using pairing. In: Information technology: coding and computing. ITCC, 1(4–6, 631–5
8. Yi L, Bai G, & Xiao G. (2000). A new type of proxy signature scheme.Electron Lett, 36(6) 527–8.
9. Sunitha and Amberker: Proxy Signature Schemes for Controlled Delegation, Journal of Information Assurance and Security, 159-174 (2008).
10. Xu,Q.,Xie Q.. Improvement of ID-based proxy signature scheme without trusted PKG.[J] Computer Applications. 2008,28(12):pp.3208-3210.

Using Data Analytics to Predict Website Phishing

Adaeze Nwaigwe[1]
Paseka School of Business, Minnesota State University, Moorhead, MN, USA

***Summary*—** The proliferation of web services delivered over the internet and serving our financial and health needs, mean that Internet users are even now, more susceptible to different threats which manifest via the web. These threats present majors risks including financial loss, identity theft, and reputation loss for businesses when the risks are realized. While no single solution exists to mitigate the risk of web phishing, identifying the factors that most likely predict a phishing attack is a major step towards creating strategies to mitigate risk. In this research, a simple decision tree algorithm is used to determine factors that most likely indicate a phishing attack. Decision trees produce more transparent results compared to the more complex black-box methods, like neural networks. This paper illustrates, that the model accuracy of the decision tree model competes strongly with that of the neural network model in previous studies. Particularly, results reveal that, the characters contained in the domain name of a website, the characteristics of the web domain's anchor, and whether or not a popup window is used to collect a user's credentials, are critical factors for classifying phishing websites.

I. INTRODUCTION

The Internet has come to bear heavily in our lives both for business use and personal use. Brick and mortar strategy for large retailers pose heavy overhead costs. Retailers are adopting omni-channel strategy to stay competitive, combining in-store shopping with online shopping and then use nearby stores as hubs to route goods to customers and perform more personalized services. Internet retailers have no stores and some retailers in this category have achieved major market shares. This trend allows consumers shop at their convenience but the strategy heightens the risk of website phishing attacks. Phishing employs social engineering and technical trickery to steal an individual's personal identity and financial account credentials [1]. Social engineering methods utilize spoofed e-mails which appear to originate from legitimate businesses and agencies. Unsuspecting individuals are then misled to bogus websites where they may disclose information like usernames and passwords. Technical deceptive methods implant malware on computing devices to criminally obtain user credentials and to redirect users to phishing websites where they would inadvertently disclose personal and financial information to unauthorized parties. Such information residing in the hands of such persons, puts its owners at risk of identity theft, financial loss and reputation loss.

Some solutions exist for thwarting phishing attacks. For instance, legal actions against the perpetrator can be undertaken. However, phishers are hard to trace and further, since phishing websites are short-lived, law enforcement agents need to act quickly in such situations. The latter might often not be feasible. Educating users of information systems is also a good solution. But,

[1] Adaeze Nwaigwe is the corresponding author.

phishing methods can be quite sophisticated, and so, this solution is not always effective [2]. There are also commercial list-based technical solutions like use of blacklist strategy. Heuristic-based methods which are feature-based, provide another solution. The feature-based solution uses a set of features to classify phishing websites.

This work supports automated feature-based solutions for identifying phishing websites. It looks to see, if a simple decision tree algorithm can identify phishing websites as effectively as more complex machine learning algorithms. This research, further seeks to determine, the factors, which are most indicative of phishing websites. Using a decision tree algorithm provides transparency as results are human readable and the impact, contributed by each factor in the model is visible.

In section II, the problem addressed by this work is presented, as well as existing research in the area. Section III describes the methodology used, what is accomplished and the results obtained. Section IV presents concluding comments and opportunities for future work.

II. RELATED WORK

Supervised machine learning is that branch of computer science which tasks machines with inferring a function from labeled training data, that is, data that includes a class or category for each observation. The inferred function is then used for mapping new examples. In the ideal case, the algorithm will correctly assign all unseen examples into classes. Supervised machine learning have successfully been used in creating a model that maps new examples to classes [3] [4] [5].

Phishing attacks can be detected by assessing the characteristics of a webpage or its web address. Rami et al. have shown that self-structuring neural networks can use the features of a website to classify it as phishing [3]. While neural nets are adaptive, they have to be retrained frequently using fresh data so as to remain effective. Further, neural network algorithms employ a black box approach, making it difficult to clearly see the impact of each feature in the inferred model. Moghini and Varjani presented a complex approach using Support Vector Machines (SVM) and a rule-based classifier to detect phishing attacks in the banking sector. Their experiment yielded an accuracy of 98.65%, kappa statistic of 0.969 and Sensitivity rate of 0.9914 [4]. Random forests have been used to identify phishing websites also, and yielded an accuracy of 99.95% [5]. However, unlike decision trees, the random forests approach, also yields results which are more difficult to interpret. In this paper, a simple decision tree algorithm, C5.0 [6], is used to classify phishing websites and to seek the most important features to consider in the process. Unlike in the previous study done in the banking domain [4], the method used in this paper is generalizable to multiple domains. Decision trees generated by C5.0 are human readable and so, easier to understand and deploy than the models generated by more sophisticated methods like neural networks and SVMs. Further, C5.0 decision tree algorithm does well on most problems and is quite efficient. Thus, the algorithm is a good choice for creating a model that will predict features that most characterize a phishing website. The seventeen features used in this work are those previously proposed by Rami et al [7].

III. RESEARCH APPROACH AND METHODOLOGY

Data used in this work was donated to the repository of the Machine Learning Institute of the University of California Irvine (UCI) [8] by Mohammed et al. [3]. The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms. The dataset originally contained thirty predictor features and a target. Of these, the original seventeen predictor features used in the study of Rami et al. were selected and used to build a decision tree

model. This was to support a comparison of the results obtained, to those in the previous study which applied an artificial neural network [3].

The seventeen features considered were: substituting an IP address for domain name, X1; using a lengthy website URL to hide the suspicious part of a URL address, X2; using "@" within a URL, X4; whether the '-' symbol is contained in the domain name of website, X6; count of dots in the domain part, X7; position of HTTPS token in URL, X12; whether request for objects like images, are loaded from a different domain, X13; whether the domain of the anchor, "<a>" differs from that of website, X14; whether the domain name of Server Form Handler (SFH) differs from that of the website, X16; similarity of Hostname in URL to websites' WHOIS identity, X18; whether website is Redirected, X19; using JavaScript to display fake URL, X20; whether right click is disabled, X21; using popup window to collect users credentials, X22; age of the domain, X24; DNS Record accessibility, X25; amount of web traffic, X26.

The features X1, X4, X6, X12, X13, X18, X20, X21, X22, X24, X25 each contained two levels, Legitimate (L), or Phishing (P). The features X2, X7, X14, X16, and X26 each had three levels Legitimate (L), Suspicious (S), or Phishing (P). X19 contained only the levels Legitimate (L), or Suspicious (S).

Of the 11,055 observations in the dataset, training and test datasets were randomly selected using a 75:25 ratio, respectively. Ten-fold cross validation was used for resampling and tuning the model. C5.0 decision tree algorithm was then used to build a decision tree model within the R environment [8]. Results are presented below.

trials	Accuracy	Kappa
1	0.9093104	0.8151194
5	0.9053265	0.8072407
10	0.9111164	0.8187844
15	0.9130472	0.8229770
20	0.9128058	0.8223606
25	0.9142536	0.8253872
30	0.9150977	0.8270515
35	0.9164245	0.8296968

Figure 1. Trial metrics for training data

(a)	(b)	<-classified as
4422	211	(a): class L
379	3280	(b): class P

Figure 2. Confusion matrix for training dataset

Evaluation on training data (8292 cases):

Trial	Decision Tree Size	Errors
0	94	624(7.5%)
1	37	1317(15.9%)
2	40	1203(14.5%)
3	51	1200(14.5%)
4	51	1165(14.0%)
5	45	1130(13.6%)
6	52	1133(13.7%)
7	50	981(11.8%)
8	49	1570(18.9%)
9	49	1159(14.0%)
10	43	1446(17.4%)
11	44	1170(14.1%)
12	53	1409(17.0%)
13	50	1213(14.6%)
14	31	1174(14.2%)
15	57	1384(16.7%)
16	38	1044(12.6%)
17	31	1473(17.8%)
18	40	1062(12.8%)
19	43	1118(13.5%)
20	57	957(11.5%)
21	51	918(11.1%)
22	57	801(9.7%)
23	68	848(10.2%)
24	53	771(9.3%)
boost		590(7.1%) <<

Figure 3. Accuracy per trial

Attribute usage:

100.00% X6P
100.00% X14P
100.00% X22P
87.86% X13P
87.82% X7S
78.88% X26S
77.69% X18P
72.24% X1P
72.18% X12P
71.68% X7P
65.46% X16P
65.23% X19S
64.44% X26P
63.90% X2P
62.40% X20P
62.05% X4P
61.03% X14S
60.38% X25P
59.58% X24P
57.62% X16S
53.69% X21P
36.28% X2S

Figure 4. Contribution of individual features to the classifier

Total Observations in Table: 2763

actual values	predicted values L	P	Row Total
L	1445	79	1524
	0.523	0.029	
P	142	1097	1239
	0.051	0.397	
Column Total	1587	1176	2763

Figure 5. Confusion matrix for test data

IV. CONCLUSIONS AND FUTURE WORK

Figure 1 shows the accuracy and kappa scores for different trials during model building. The final C5.0 classification model selected was characterized by a model accuracy of 92.9% as indicated by the confusion matrix of figure 2 and optimized at only 25 trials as shown in figure 3. Performance of the model on test data resulted in model accuracy of 92% (figure 5). It must be noted, that the results obtained in this work are comparable to those obtained using a neural network model which yielded a training set accuracy of 94.07%, a validation set accuracy of 91.31%, a testing set accuracy of 92.18% at 1000 epochs of training [3].

Figure 4 shows that, for features X6, X14 and X22, their phishing (P) level was used in classifying 100% of the training cases. This means that the P level of these three features are key to identifying phishing websites. X13P, X7S, X26S, X18P, X1P, X12P, and X7P contributed over 70%. X14S was used 61% of the time.

This research illustrates that decision trees, which produce more transparent results can yield a classification model with accuracy comparable to that produced by the more complex neural network method, when tasked with identifying phishing websites. Comparison of algorithm efficiency in terms of time to create models, as well as its performance when more features are considered, would further help choose the better algorithm.

References

[1] APWG, "Phising Activity Trends Report, 1st Half 2017," 17 October 2017. [Online]. Available: http://docs.apwg.org/reports/apwg_trends_report_h1_2017.pdf. [Accessed 26 January 2018].

[2] R. Dhamija, J. D. Tygar and M. Hearst, "Why Phishing Works," in *CHI 2006 Conference on Human Factors in Computing Systems*, Montreal , 2006.

[3] R. Mohammad, F. A. Thabtah and T. L. McCluskey, "Predicting Phising Websites Based on Self-Structuring Neural Networks," *Neural Computing and Applications,* vol. 25, no. 2, pp. 443-458.

[4] M. Mohmood and A. Y. Varjani, "New Rule-Based Phising Detection Method," *Expert Systems with Applications,* vol. 53, pp. 231-242, 2016.

[5] R. S. Rao and A. R. Pais, "Detection of phishing websites using an efficient feature-based machine learning Framework," *Neural Computing and Applications,* 2018.

[6] RULEQUEST RESEARCH, "C5.0: An Informal Tutorial," March 2017. [Online]. Available: http://www.rulequest.com/see5-unix.html. [Accessed 26 January 2018].

[7] R. Mohammad, T. L. McCluskey and A. Fadi, "An Assessment of Features Related to Phishing Websites using an Automated Technique," in *International Conferece For Internet Technology And Secured Transactions. ICITST 2012*, London, 2012.

[8] UCI, "Phishing Websites Data Set," UCI Center for Machine Learning and Intelligent Systems, [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Phishing+Websites. [Accessed 26 January 2018].

[9] T. R. Foundation, "What is R?," [Online]. Available: https://www.r-project.org/about.html. [Accessed 26 January 2018].

System Security Vulnerability

Xin Fang Mak
Robert Morris University, Moon Township, Pennsylvania, USA

***Summary*—**In 2016 alone, over 50% of the disclosed software defects were remotely exploitable while 18 percent had a severity rating of 9 or higher on a scale of 1 to 10[1]. The numbers are forecasted to increase substantially, and it was suggested that 2017 was the worst year on record for system security vulnerabilities. The increase of system security vulnerabilities has furthered the need for safety measures and awareness. Research shows that as many as 85% of targeted attacks are avoidable. [1] This paper reviews the primary causes behind system security vulnerabilities and anticipated solutions. It is hoped that this study will inform users and IT professionals about the severity of simple system security vulnerabilities and how to counter them.

I. INTRODUCTION

The European Union Agency for Network and Information Security (ENISA) defines vulnerability as the 'The existence of weakness, design or implementation error that can lead to an unexpected, undesirable event, compromising the security of the computer system, network, application, or protocol involved.' [2] This paper highlights seven-major categories of system security vulnerabilities.

With the constant and rapid change in technologies, users need to be alert and proactive to avoid potential threats. A small vulnerability in a system could potentially impact not only a business's reputation but could also cause a financial blow to the organization. As attackers get savvier, they find ways to exploit existing vulnerabilities and can even launch simultaneous attacks against numerous systems.

II. VULNERABILITY SURVEY

System vulnerability can come in the form of unauthorized access or malicious behavior such as viruses, worms, Trojan horses and other types of malware. These vulnerabilities can result from software bugs, weak passwords or software that has been infected by a computer virus or script code injection. New patches or fixes need to be applied to avoid the potential for hackers and malware to compromise system integrity. Software security updates are important as the patches can remedy flaws and security loopholes from the initial release. [3]

There are different classifications of vulnerabilities which include hardware, software, network, personnel, physical site, and organizational systems. [4] Examples of the commonly found vulnerabilities are buffer overflows, invalidated input, race conditions, access-control problems and weakness in authentication, authorization and cryptographic practices.[2]

Seven fundamental vulnerability causes are as follows. [5] The first cause is system complexity. A complex system increases the probability of flaws and unintended access points.

[1] National Vulnerability Database

[2] Apple Developer Guide 2016

The next cause is familiarity. With standard software, operating systems and hardware on the market, the likelihood of attackers to exploit known flaws increases. Many of the vulnerabilities discovered in IPv4 protocol software for example, were also found in the new IPv6 implementations. The third cause is connectivity based. Many devices connect to multiple networks and each connection point increases access and vulnerability for potential attackers. The next cause is password management. Users tend to use weak passwords that can easily be discovered by brute force and commonly keep passwords written in plain sight. The fifth cause is operating system configuration where the user chooses to enforce suboptimal policies on user or program management such as internet browsing. It is common to find websites containing spyware or adware that can be automatically installed on the user's devices. Some malicious software collects information before that is sold to thirds parties. An enabled firewall can help prevent some of this malware if it isn't disabled. The sixth cause is software bugs. A programmer can unintentionally leave an exploitable bug in a software program which an attacker can exploit to misuse an application. The seventh cause is unchecked user input, such as buffer overflow, SQL injection or other non-validated data.

III. PROPOSED SOLUTIONS

1. Encrypted file system for sensitive data

Implementation of an encrypted file system for raw data is essential, especially when the system contains valuable information. This acts as an additional layer of security in the event that critical data is obtained by unauthorized individuals or malicious software. Without proper decryption methods, this critical data will remain safe, even in unsavory hands.

2. Authentication and Security Measures

Strong and mixed passwords should be used and changed frequently even with multiple point of authentication and security measures. This includes VPN, simultaneous data and voice (SDV), and Wi-Fi passwords. In an organizational setting, authentication and access levels are a must for employees to access sensitive data. For example, an average employee should only have access to systems and files, that are needed to fulfill his/her duties. Any special requests should always be sent to a team for approval, not just the person who can authorize the request.

3. Access Control and Policies

The third suggested solution is the implementation and enforcement of device access control, device policies regarding the eligibility of the types of devices, and accessibility time. It is also important to keep the policies up to date. [7] For example, in 2008, the Department of Defense developed policies that banned USB and other removable media from entering/exiting their environments. According to NIST, 'a state of access control is safe if no permissions can be leaked to an unauthorized or uninvited principal.' To guarantee the effectiveness of an access control system, it is paramount that it doesn't result in unauthorized access.

4. Testing (Penetration & Vulnerability)

From time to time, it is essential for organizations to perform a vulnerability assessment and penetration testing to understand their overall security risk, but when, and which testing should be

conducted? Testing can provide benefit to any information program, and it is a fundamental component of a Threat and Vulnerability Management Process.[3] [Secureworks, April 2015]

Vulnerability Assessment or Audit

The goal of a vulnerability assessment is to attain a prioritized list of vulnerabilities in the environment so that remediation can occur. It focuses on breadth over depth of the vulnerabilities that the system is expose. [7] This assessment is beneficial for organizations that need help identifying potential issues in their systems. [Miessler, D. Jan 2014]

Penetration Testing

Commonly known as a 'Pen Test', this test is used to determine whether the system can withstand an intrusion attempt from an advanced attacker. [8] Unlike vulnerability assessment, pen testing focuses on depth as opposed to breadth. SecureWorks divides the testing into two types, white box and black box testing. White box testing uses a vulnerability assessment and other pre-disclosed information to test for penetration points, while black box testing performs with limited knowledge of the target systems prior to testing.

IV. CONCLUSION

Vulnerabilities have been found in every major operating system such as Windows, MacOS, and various forms of Linux. One of the ways to reduce the probability of a vulnerability is through constant vigilance, including careful system maintenance (applying updated software patches) [9], best practices in deployment (the use of firewalls and access controls) and auditing (during deployment and throughout the deployment lifecycle). The fundamental concept of information security is the principle of defense in breadth and depth. Seven unique types of vulnerabilities are system complexity, system over standardization, over-connectivity, password management, operating system configuration, software bugs, and unchecked user input. Setting up a multi-layer security policy that can prevent exploitation, detect and intercept attacks is crucial. Two diagnostic tools that can help institute a strong security system are a vulnerability assessment and a penetration test. Using these to improve all systems greatly reduces vulnerability to cyber-attacks.

[3] Secureworks is a subsidiary of Dell Technologies that provides information security services.

References

[1] Vijayan, J. (2017, May 23). Data Breach, Vulnerability Data on Track to Set New Records in 2017. Retrieved July 31, 2017, from https://www.darkreading.com/attacks-breaches/data-breach-vulnerability-data-on-track-to-set-new-records-in-2017/d/d-id/1328947

[2] 'Glossary' *Glossary - ENISA*, European Union Agency for Network and Information Security, 20 Jan. 2016, www.enisa.europa.eu/topics/threat-risk-management/risk-management/current-risk/risk-management-inventory/glossary.

[3] What is Vulnerability? - Definition from Techopedia. (n.d.). Retrieved July 31, 2017, From https://www.techopedia.com/definition/13484/vulnerability

[4] Pananini, P. (2016, February 23). The Top Five Cyber Security Vulnerabilities. Retrieved July 31, 2017, from http://resources.infosecinstitute.com/the-top-five-cyber- security-vulnerabilities- in-terms- of-potential-for- catastrophic-damage/

[5] Overview of Cyber Vulnerabilities. (n.d.).Retrieved July 31, 2017, from https://ics-cert.us-cert.gov/content/overview- cyber-vulnerabilities.

[6] Harrison, M. A., Ruzzo, W. L., & Ullman,J. D. (2015, May 6). Access Control Policy and Implementation Guides. Retrieved July 31,2017, from http://csrc.nist.gov/projects/ac-policy-igs/index.html

[7] System Vulnerability Assessment. (n.d.). Retrieved July 31, 2017, fromhttp://scwoa.com/services/system-vulnerability-assessment/

[8] Drew, S. (2015, April 8). Vulnerability Assessments versus Penetration Tests.Retrieved July 31, 2017, fromhttps://www.secureworks.com/blog/vulnerability-assessments- versus-penetration- tests

[9] Secure Coding Guide. (2016, September13). Retrieved July 31, 2017, from https://developer.apple.com/library/content/documentation/Security/Conceptual/SecureCodingGuide/Articles/TypesSecVuln.html

Advocating for Logical & Physical Network Segmentation to Reduce Personal Health Information (PHI) Disclosure

Cody R. Crawford
Department of Cyber Security, University of Maryland University College, Largo, MD, USA

***Summary*—** The failure of Patch Management in hospital administrations directly led to their infection by the infamous "Wannacry" Malware on 12 May 2017 [1]. Hospitals are often at risk of infection by ransomware due to a combination of factors. These include the criticality of uptime, poor allocation of maintenance resources, ineffective patch management, and a lack of emphasis on cybersecurity by manufacturers and hospital administrations [2]. This paper advocates for a simpler solution to this problem; Logical and Physical Network Segmentation. It discusses the creation of two networks. One for use in day-to-day business & connected to the Internet, and one another to process Personal Health Information (PHI) connected to the rest of the hospital in an Intranet. By separating the two networks we can minimize the risk of PHI compromise by unsophisticated malware such as Wannacry; for which patches existed prior to the main infections. This method will greatly reduce the attack surface by focusing protection onto the intranet protecting PHI, therefore allowing for more efficient utilization of resources to protect PHI. This protection method does not replace the need for robust identity management systems, access control, and other technical measures to mitigate Zero-Day Exploits.

I. INTRODUCTION

Cybersecurity threats are numerous and unrelenting. The Sophos 2018 Malware Report demonstrated that Wannacry was responsible for an estimated 45.3% of all infections during 2017, many of them in hospitals, Industrial Control Systems (ICS), or personal machines [3]. The infections primarily were an issue of Patch Management, or the implementation of cybersecurity programs/policies [1]. The infection vector was via the Windows Server Message Block (SMB), a patch for which existed a few months prior to the Wannacry infections [3].

Healthcare systems are valuable commodities, and need to be protected as such. The need for access to records can be the difference between life and death, so hospitals will sometimes pay the ransom. In one case Hollywood Presbyterian payed $17,000 for the return of its lost information [3]. As these attacks become commonplace, it will become more fiscally beneficial to implement more stringent cybersecurity measures.

The Cybersecurity community needs to advocate for secure solutions. This paper proposes Logical and Physical Network Segmentation to protect critical data and systems. In our current "Internet of Things" Age, we need to make wise decisions on what needs a network connection, and what does not [4]. A toaster does not need to be connected to the Internet, but hospital systems may fit that requirement. The paper is organized as follows: the proposed network topology is shown in Section II, demonstrating the use of Logical and Physical Network Segmentation. Section III discusses the continued need for Patch Management and robust cybersecurity policy in the face

of determined attackers. These could include rogue employees or hackers using Zero-Day Exploits. Concluding remarks are presented in Section IV.

II. PROPOSED NETWORK TOPOLOGY

One of the most important systems in a hospital is a standard computer. It is through these systems that physicians, clinicians, and nursing staff update patient medical records, schedule surgeries, research symptoms, and ultimately provide potentially life-saving treatment. The information protection needs for these systems are vast, and not always deployed in a manner commensurate with the need for security. Instead, a hospital operates more on the "Availability" side of the Confidentiality-Integrity-Availability Triad [5]. Information protection is important, but dwarfed by the operational requirement to provide life-saving care. Network segmentation is a general template for security that is costly up-front, but useful in the long-term.

In the proposed model, all PHI should be stored and processed on an intranet, a local or restricted communications network. All PHI disclosure should be handled by specific machines within the hospital network, which are kept patched and use the latest software, operating systems, etc. Only those PHI disclosure systems would access both the Intranet and the Internet; they are gateways. Access controls must be placed via firewall, or on external network boundary switches. These should block all outbound attempts except from specific machines. Access to the Intranet from user devices would be done via remote-login software to a central server that runs the Intranet. Most systems are physically separated via the Open Systems Interconnection model, but connected at the Application level. An example diagram has been generated below.

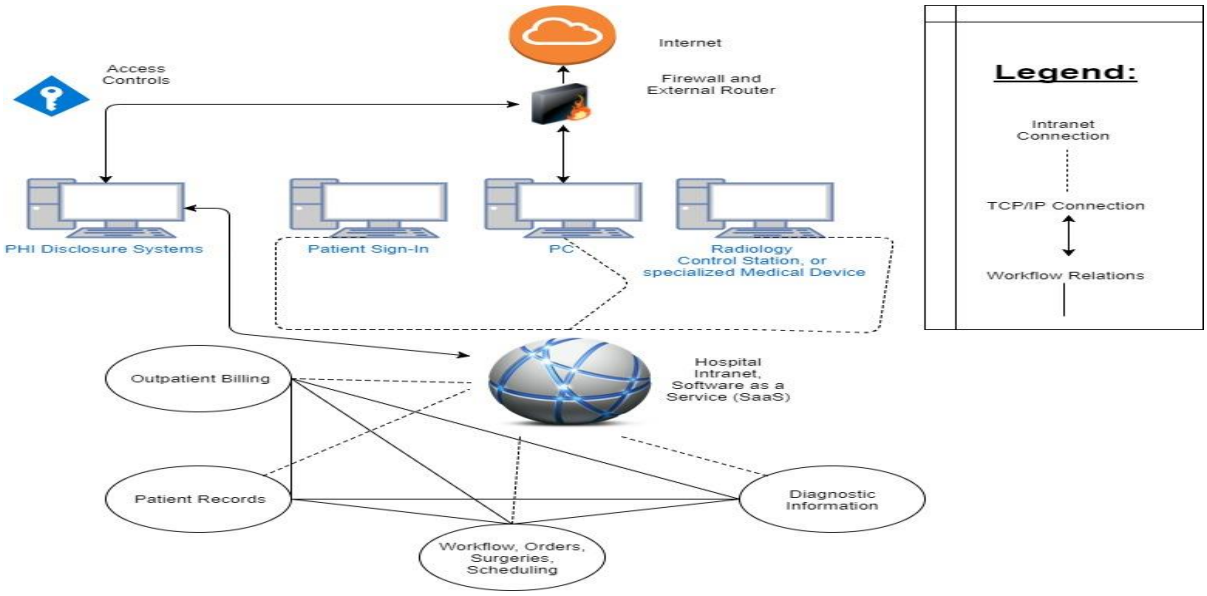

Figure 1

Software controls would be placed to limit involuntary disclosure. Measures should be implemented that block inter-process communication or transfer at the Application Layer unless initiated by the user in conjunction with a hard-ware token such as a SmartCard. The ability to "copy and paste" data from inside the remote environment to the outside should not exist. TCP/IP would be enabled only on PHI release machines in a specific subnet or department, and on user workstations. Doctors, clinicians, and other personnel should be able to log into the Intranet via their ID Card, acting as a hard-ware token for use with 2-Factor Authentication (2FA).

On the infrastructure side; point-to point encryption should be enabled. This would be done via the use of an appropriate algorithm and peer-reviewed implementation, such as AES-256. The main vulnerability in the proposed topology comes from the sniffing of wireless traffic between machines or side-channel attacks [5].

The benefits of this method are two-fold. First, all PHI is encrypted and decrypted between machines (local software) vs. logical devices that allow tampering. Second, PHI loss is prevented at the Network Layer unless authorized via the PHI disclosure office. This heavily narrows the presented attack surface, allowing our Hospital to focus its efforts more easily. It also limits the scope of investigations into PHI disclosure, reducing the time needed to detect a malicious Insider Threat.

In general, the proposed hospital network is separated via Intranet, and connected via software within a virtualized environment. Billing, PHI, Workflows, Scheduling, and Diagnostics are all processed via a central server that hosts the intranet. For information to flow outside of this network, it must go through the PHI disclosure systems. PHI is encrypted-at-rest, and during transit. Access Controls are present on endpoint devices, such as the external router or firewall, that restrict PHI disclosure. Production systems reside on the Intranet. Standard PC's/Desktops have access to the Internet via TCP/IP for research, email, or other activities. Hospital personnel should be trained and inspected quarterly to ensure they are not putting PHI outside the Intranet.

III. THE NEED FOR PATCH MANAGEMENT & ROBUST CYBER POLICY

The policies and Network design explained previously rely on specific disclosure systems for the hospital to function. The PHI disclosure systems must therefore be maintained and staffed for "surges"; mass-casualty events or days where patient records are requested on an emergency basis. That is not to say they must be based entirely on manpower; only that they should avoid replacement as "another boundary technical solution". These systems, acting as the only liaison between the hospital intranet and production work-center (Surgery, Radiology, etc), must be kept up-to date with the latest patches and security software. Patch Management is a must; hot-swappable machines should also exist on standby for any contingency events not otherwise planned for.

Robust policy must also exist to prevent disclosure via other methods. Most organizations are focused on external threats, to include hackers, hacktivists, and automated intrusion systems. They should also be ready to detect network compromise via trusted insiders from new hires, disgruntled employees, or other malicious actors with physical access to systems. Desired information could include PHI such as a political official's sexual history or STD records, surgeries like abortions or or Vasectomies, all could be used for blackmail. A robust Identification Management System should therefore exist; followed by weekly or monthly audits and inspections.

The ID management system serves two purposes. First, it is typically coupled with a hard-ware token, such as a Smartcard, required for logging onto the Intranet. The other purpose is to log actions on the intranet, detailing what records were pulled, when, at what machine, where they were sent, etc. If an employee's computer is breached, the hacker should be unable to talk to the Intranet as they would lack the hard-ware token to establish a connection. Other cybersecurity precautions should therefore be enacted catch instances of compromise and quarantine those machines.

Mobile devices should be minimized as much as possible, with few exceptions. Doctors should not be working after-hours on PHI at home. To do so would require management of a

Virtual Private Network- which complicates cybersecurity. Instead they should take patient charts home with them and be responsible for the records via a sign-out policy. We should encourage the use of mobile tablets within the hospital premises; provided they enable 2FA via Smartcard readers. A determined attacker would then require physical access to the device, a password/code, and the owner's hard-ware token. This limits the list of potential insiders when conducting an audit, or a post-incident forensic response. It also expands the operational capability of life-saving care, with a small impact to the hospital's cybersecurity posture.

IV. CONCLUSIONS AND FUTURE WORK

The primary issue of the day is changing the collective mindset about cybersecurity. It absolutely must be thought of in risk management terms; e.g "If we choose not to enforce this policy, what is our potential loss long-term vs. the up-front investment"? Until businesses and healthcare organizations begin to think in these terms, we will not see real progress. It is possible to have patched and secured networks, but only if the right personnel are hired and the requisite strategic guidance is carried out or enforced up and down the leadership chain.

In the interim, establishing segmented networks at the logical and physical levels will mitigate most malware out in the wild. Robust cybersecurity policies and actions are still needed to tackle Zero-Day Exploits. If we do nothing; then we accept the risk of continued PHI exploitation. This is counter to the Hippocratic oath; which guides the medical profession [6]. The status quo is a violation of HIPAA intent [7]; with little in the way of mitigation being carried out. We must do better, and develop models that achieve both secure communications capability and enable life-saving care.

References

[1] Boiten, E., & Wall, D. S. (2017, October 31). WannaCry Report Shows NHS Chiefs Knew of Security Danger, but Management Took No Action. Retrieved January 02, 2018, from https://www.scientificamerican.com/article/wannacry-report-shows-nhs-chiefs-knew-of-security-danger-but-management-took-no-action/

[2] Olenik, D. (2017, November 01). WannaCry, Cerber most used ransomware types, hospitals most hit sector, report. Retrieved January 02, 2018, from https://www.scmagazine.com/wannacry-cerber-most-used-ransomware-types-hospitals-most-hit-sector-report/article/704746/

[3] Sophos Labs. (2017, October 30). Sophos Labs 2018 Malware Forecast. Retrieved January 2, 2018, from https://media.scmagazine.com/documents/321/sophos_2018_malware_forecast,__80124.pdf

[4] Morgan, Jacob. "A Simple Explanation Of 'The Internet Of Things'." Forbes, Forbes Magazine, 20 Apr. 2017, www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#cc6efcf1d091.

[5] Platt, W. (2009). Wireless at the hospital and the threats they face. Information Security Reading Room. Retrieved July 10, 2017, from https://www.sans.org/reading-room/whitepapers/wireless/wireless-hospital-threats-face-33003

[6] Lasagna, L. (n.d.). Guides: Bioethics: Hippocratic Oath, Modern version. Retrieved July 10, 2017, from http://guides.library.jhu.edu/c.php?g=202502&p=1335759

[7] Rousse, M. (2016, December). What is HIPAA Privacy Rule? - Definition from WhatIs.com. Retrieved July 10, 2017, from http://searchhealthit.techtarget.com/definition/HIPAA-Privacy-Rule

Autopsy: An Overview

Julia R. Balzano
Department of Computer Information Systems, Robert Morris University, Moon, PA, USA

Abstract: Autopsy is a digital forensics tool created by the Sleuth Kit to aid in the acquisition and analysis of digital evidence. Digital evidence differs from physical evidence in that it has a wider scope in the types of evidence that can be discovered, can comprise of physical and/or personal information, and involves criminal issues that supersede law enforcement's usual role in evidence collection. Autopsy's latest iteration, Autopsy 3, contains an array of features to analyze evidence, an auto-generated documentation system, and an intuitive design. The cost-effectiveness makes Autopsy an appealing choice for investigators. The software has some limitations, however, such as its usability and its discovered evidence'possible inadmissibility in a court of law when the software is used without further validation from another digital forensics software. Further study into various mobile devices, wearables, and data from the Cloud provides areas for Autopsy to grow and develop.

Keywords: Autopsy, The Sleuth Kit, Digital Forensics, Digital Evidence

I. **INTRODUCTION**

This paper examines the digital forensics tool Autopsy, a graphical interface to the Sleuth Kit, and its role in acquiring, discovering, and analyzing various types of electronic evidence. Autopsy is a free software supported by Basis Technology, a leading developer in digital forensics products
and multilingual text analytics, and is designed to provide an intuitive tool for file system image analysis without sacrificing any
depth. Autopsy comes packaged with an abundance of features.

In today's modern technological landscape, the need for digital forensics examination tools is greater than ever. Modern society is characterized by an abundance of gadgets, most of which can be found in pockets and in homes. The rise in popularity in other technologies such as information repositories and network traffic also makes forensics cases ever-growing in complexity [1]. In the past few decades, U.S. courts have allowed digital evidence retrieved from ATM transaction logs, word processing documents, spreadsheets, and even logs from a hotel's electronic door locks to be admitted [2] Most people, regardless of their willingness, leave behind some sort of digital footprint, whether it is information found on a smart phone or GPS data found in the computers inside cars. In the field of solving crimes and preparing court cases, finding digital evidence can be crucial for law enforcement officials and prosecutors alike.

Due to the prevalence and potential wealth of information found in digital evidence, using the proper tools to examine this data is a critical part of an investigation. Autopsy is a trusted forensic tool that is used by law enforcement, military, and even corporate examiners [4]. In addition to its incredibly fast speed and the fact that the software is free, Autopsy is capable of examining a variety of devices that may hold digital evidence. First, Autopsy can find and examine information acquired from a personal computer. This information can include data from Internet browsing, such as programs that maintain temporary Internet files, cookies, and browser history, as well as emails and downloaded files or applications [5]. However, Autopsy

can also analyze information brought over from portable electronic devices, such as cell phones. This is imperative because, currently, the primary focus of interest to examiners and researchers has become processing digital evidence from portable electronics [5]. Autopsy can even recover photographs from a camera's memory card [4]. Finally, Autopsy was designed to be a user-friendly tool, both in its design and the resources available to users. An online user documentation guide provides instructions on installation, creating cases, importing data, and a framework to Autopsy's various modules, some of which come from third-party developers. For its more technologically-savvy users, Autopsy also provides a developer's guide which instructs users on how to develop their own modules. These factors contribute to Autopsy's popularity among both law enforcement agencies and private examiners.

II. TECHNOLOGY

Autopsy's technology has been developing since March 2001, when it was developed as a graphical interface to another Sleuth Kit tool, The Coroner's Toolkit (TCT) and was compatible only with Linux and OS X [6]. Most of the code in Autopsy 1 and 2 was developed by Brian Carrier, while Autopsy 3 was largely built and financed by Basis Technology. Autopsy 3 began development in 2010 from scratch. It drew its base design from discussions held at the Open Source Digital Forensics Conference, in which developer Brian Carrier was a key speaker. Version three of Autopsy expanded from the previous versions'exclusivity to Linux and OS X, and was Windows-based and automated [6]. Although Basis Technology provided the main funding for version three, other funding was obtained from the U.S. Army and 42Six Solutions, a company of technology-based engineers with a focus on interface-driven development. In September of 2012, Autopsy 3 was released and remains the most current version of the software.

Autopsy 3 focuses on three key points in its development and subsequent technologies: extensibility, ease of use, and speed of results. Its extensibility was a design choice that had an intense focus during development. Autopsy 3 was designed to be an end-to-end platform with both pre-programmed modules and others available from third-party developers [7]. The list of features already contained within version three include many technologies that increase intuitiveness, ease-of-use, and the options for analyzing data. The Multi-User Cases feature allows many examiners to work on a single, large case at a time [7]. Time Analysis aids in organization by displaying system events in a graphical interface in order to identify activities [7]. Another feature is Keyword Search, which can aid in examining large forensic images by narrowing searches to specific terms and regular expression patterns. To accomplish this, the feature utilizes text extraction and index searched modules [7]. Web Artifacts is another feature contained within version three that can extract web activity from common browsers in order to assist in identifying user activity [7]. There are three analysis tools in Autopsy 3: Registry Analysis, LNK File Analysis, and Email Analysis. Registry Analysis uses RegRipper, an open source tool written in Perl for the extraction and parsing of data, to identify recently accessed documents and USB devices [7]. The LNK File Analysis feature can identify short cuts and accessed documents, while Email Analysis can parse MBOX format messages, such as Thunderbird [7]. To extract geographic location data and camera information from JPEG files, Autopsy 3 utilizes its EXIF feature. Other technologies included in Autopsy include File Type Sorting, Media Playback, Tags, and a Thumbnail viewer in order to group like files, tag files and add comments, and view photos and videos from within the software. The Robust File System

Analysis feature supports several file systems, such as NTFS, FAT12/FAT16/FAT32/ExFAT, HFS+, ISO9660 (CD-ROM), Ext2/Ext3/Ext4/, Yaffs2, and UFS from The Sleuth Kit [7]. Autopsy 3 also uses Hash Set Filtering to filter out known good files using NSRL and flags known bad files using custom hashsets in HashKeeper, md5sum and EnCase formats [7]. Another of Autopsy 3's features extracts strings from unallocated space and unknown file types [7]. Titled Unicode Strings Extraction, this feature can extract strings in many languages such as Arabic, Chinese, and Japanese. Further Autopsy 3 features include File Type Detection, Interesting Files Module, and Android Support, which can extract data from SMS, call logs, contacts, Tango, and Words with Friends.

Autopsy can analyze disk images, local drives, or a folder of local files [7]. Disk images can be formatted in either raw/dd or E01 format [7].

In a digital forensics examination, documentation is one of the most critical components, used to establish a concrete timeline of evidence and the investigators involved. Autopsy supports this need by providing an extensible reporting infrastructure, with HTML, XLS, and Body file reports available by default [7]. Each type of report is customizable by the investigator in order to pick and choose the types of information to be included. The HTML and Excel reports are intended to be fully packaged and shareable, and may include references to tagged files along with any comments and notes added by the investigator as well as other automatic searches performed by Autopsy during ingest [7]. These automatic searches include bookmarks, web history, recent documents, keyword hits, hashset hits, installed programs, devices attached, cookies, downloads, and search queries [7]. The other report type, Body File, can be utilized primarily for examining timelines. The Body File will include MAC times for every file in an XML format for import by external tools, such as mactime in the Sleuth Kit [7]. For reporting, investigators can generate multiple reports at a time and edit modules or create new modules to fit their their purpose.

Another key point in Autopsy's development is its intuitive design. Autopsy maintains an easy-to-use interface in order to be used in the most effective way by even non-technical investigators. The software uses wizards to help the investigator know what to do next, uses commonplace navigation techniques to aid in finding results, and automates processes as much as possible to prevent human error [8]. To aid non-technical users, Autopsy uses various features. Wizards are used in several places across the software to guide users [8]. A history of activity is maintained, enabling the user to back track after going down a path of investigation by using back and forward buttons [8]. Finally, previous settings are saved with the modules so that investigators can more easily analyze the next forensic image with the same settings used to analyze the last [8]. The default interface is designed simply, with the results of an analysis found in a tree on the left side of the screen. This eliminates the need to search through multiple layers of menus and tabs. Autopsy also is non-invasive with popups and messages from any tasks running in the background [8]. To avoid the risk of distraction, modules send result messages to an ingest inbox that can be opened and reviewed by the investigator.

An important key point in Autopsy's technology is the ability to produce fast results. Several features allow evidence to be processed and collected. Multiple ingest modules run in parallel to effectively utilize multi-core systems and user folders and files are prioritized over system folders and files [9]. Time intensive steps in analysis can be disabled for a faster, yet less thorough search [9]. Options such as skipping the search for orphan FAT files and skipping analysis of unallocated space are available to produce faster results. Finally, results from the

ingest modules are mostly given as soon as they are discovered. Feedback on which modules are running and what they are reporting can be found in the ingest inbox.

III. EVALUATION: PART 1 - STRENGTHS

One of Autopsy's strengths that is most appealing to both private corporations and law enforcement is its cost-effectiveness. Autopsy is a free software. This fact allows any agency or company to utilize Autopsy, regardless of budget. In comparison to another digital forensics tool, Forensic ToolKit by AccessData charges $3,995 for a perpetual license. Further, Autopsy provides the same basic features as other computer forensic tools and offers other features, such as web artifact analysis and registry analysis, that cannot be found in other commercial tools [4].

Another of Autopsy's strengths is its extensive user documentation, which can be found directly from The Sleuth Kit's website. Starting with a guide to installation, quick start, Autopsy workflow, and creating a new case and data sources, Autopsy's user documentation then dives into guides for each module. Guides for the modules are written in a way that is easy to understand and helpful for users who are just learning Autopsy. Each module guide begins with a "What Does it Do"section which details exactly what the module is and how it works. The next item listed is "Configuration,"which details if a module needs to be configured and how to do so. The last three sections are "Using the Module,"which provides instructions, "Ingest Settings,"which lets users know if they need to adjust any runtime settings, and finally "Seeing Results,"which provides an image of the module in use and the results it can provide. The user documentation also covers manual analysis features, such as the tree viewer and result viewer, reporting and tagging, and installing third party modules. Autopsy's strength in user documentation ensures that nearly any question or troubleshooting issue a user has can be found and answered without searching across the entire web.

An important strength in Autopsy's design is the number of features packaged with the software. As detailed in the technologies section, Autopsy's numerous features allow investigators to thoroughly examine a forensic image. The technological details directly contribute to Autopsy's strength. For example, the EXIF feature is a strength to investigators because, by viewing an exact geographic location and camera information, a lead as to the whereabouts of a suspect or stolen camera may be uncovered.

IV. EVALUATION: PART 2 - LIMITATIONS

While Autopsy has a number of strengths in its design, technologies, and support, it is not without limitations. Research conducted by D.J. Bennett and P. Stephens of Canterbury Christ Church University reveals a number of areas in which Autopsy can be improved for usability. As defined by Bennett and Stephens, usability in digital forensics tools refers to characteristics in the interaction between the investigator and the computer system with which they are working [10].

One limitation cited in Bennett and Stephens'research is the inconsistencies in language used in the software. Autopsy uses many terms that may not be familiar with new users of forensic tools or the Linux operating system [10]. Terms that may be outside the vocabulary of new users include 'case', 'image', 'host', 'hash database', and 'meta''and 'symlink', which are shorthand for metadata and symbolic link and may confuse those unaware of this language [10]. Bennett and Stephens also propose that new users may have difficulty differentiating between

the usage of ‘disk’and ‘partition’. Drawing from this research, Autopsy could be improved by providing clearer distinctions between terms and providing either simpler language or keys within the interface to define each term.

A limitation in Autopsy’s design that can be improved upon is its lack of depth in error prevention. An example of this can be found in the repeated use of text entry fields for entering the details of files available on the system [10]. If a user mistypes data into the system, invalid data can be introduced which can then put the entire case into jeopardy [10]. Another failing in Autopsy’s error prevention comes from its requirement to enter an investigator name, a field that requires no spaces in the text field. However, a user who does enter any spaces is not prevented from continuing into the software [10]. This error may not be discovered until much later in the case when the user cannot select an investigator [10]. This error’s cause is still not obvious, and the user would have to think back several screens in order to find out what has gone wrong.

A final limitation that lies not in Autopsy’s design but in its reputation comes from a comparison to Forensic ToolKit. Forensic ToolKit is known as a court accepted digital investigation platform, while Autopsy does not yet have this distinction, as it is an open source tool. Generally, open source tools’reliability is questioned in court due to their lack of authentication, uncertainties about chain of custody, and the validity of the acquired information, sometimes causing evidence found in such tools to be inadmissible [11].

V. CONCLUSION

The field of digital forensics is an ever-growing study that most law-enforcement, corporations, and attorneys must deal with on a daily basis [12]. As technology becomes more portable and powerful, greater amounts of information can be created, stored, and accessed [12]. With the popularity and wealth of data available on such devices, the acquirement of any possible evidence becomes an imperative task, along with the need for proper digital forensics tools. Autopsy is a tool used by corporations, government, law enforcement, and the military to accomplish this. The technology and design of Autopsy provides numerous features in the analysis of a forensic image and its reporting function ensures that all investigations are properly documented. Autopsy has some limitations in usability and possible admissibility; however, the software remains among the most popular and trusted of its kind. Future areas of study could include continued research into the analysis of images acquired from mobile devices and the Cloud, a field which is largely still being explored.

REFERENCES

[1] Harichandran, V. S., Breitinger, F., Baggili, I., &Marrington, A. (2016). A cyber forensics needs analysis survey: revisiting the domain's needs a decade later. Computers & Security 57, March 2016, pp. 1-13. doi:10.1016/j.cose.2015.10.0 [Online]

[2] Ahmed, N. (2010). Importance of Digital/electronic evidence in courts of law. [Online] Retrieved from http://www.academia.edu/20319123/_Importance_of_Digital_electronic_evidence_in_courts_of_law_.

[3] Dowling, A. (2006). Digital evidence. *University of Otago.* [Online]Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/248770.pdf.

[4] Carrier, B. (2017). Autopsy. *The Sleuth Kit.* [Online] Retrieved from https://www.sleuthkit.org/autopsy/.

[5] Law Enforcement Cyber Center. (2016). Understanding Digital Evidence. *Law Enforcement Cyber Center.* [Online]Retrieved from http://www.iacpcybercenter.org/investigators/digital-evidence/understanding-digital-evidence/.

[6] Carrier, B. (2013). Autopsy 3: Extensible Desktop Forensics by Brian Carrier. *Open Source Digital Forensics Conference.* [Online]Retrieved from https://www.slideshare.net/basistech/autopsy3-osdfcon.

[7] Carrier, B. (2017). Autopsy. *The Sleuth Kit.* [Online]Retrieved from https://www.sleuthkit.org/autopsy/features.php.

[8]Carrier, B. (2017). Autopsy. *The Sleuth Kit.* [Online]Retrieved from https://www.sleuthkit.org/autopsy/intuitive.php.

[9] Carrier, B. (2017). Autopsy. *The Sleuth Kit. [*Online]Retrieved from https://www.sleuthkit.org/autopsy/fast.php.

[10] Bennett, D. J. and Stephens, P. (2009) *A cognitive walkthrough of Autopsy Forensic Browser.* Information Management & Computer Security, 17 (1). pp. 20-29. ISSN 0968-5227. [Online]

[11] Goodison, S.E., Davis, R.C., & Jackson, B.A. (2015). Digital Evidence and U.S. Criminal Justice System. [Online] Retrieved from https://pdfs.semanticscholar.org/ee70/100415142a59e67243521819b240d2c4136c.pdf.

[12] Sykalski. (2012). Do-It-Yourself Mobile Forensics. *Southern Methodist University.* [Online]Retrieved from https://s2.smu.edu/~lsykalski/papers/MobileForensics.pdf.

Privacy and Security of Healthcare Data (A Case Study Analysis)

Hossein Zare, MS of Cyber Security, PhD; **Victoria Glazer**, Master of Cyber Security, UMUC; **Noelani Kaluhiwa**, Master of Cyber Security, UMUC;
Ian Plitt, Master of Cyber Security, UMUC; **Mojgan Azadi**, Informatics MSN, PhD.
Corresponding Author: **Hossein Zare;** University of Maryland University College (UMUC), Johns Hopkins University, Course Chair Adjunct Assistant Professor (Hossein.Zare@Faculty.UMUC.edu)

Abstract

The healthcare industry is one of the largest targets for identity theft. The information stored in electronic medical report (EMR) systems–social security numbers, names, dates of birth, and mailing addresses–make healthcare databases one of the most attractive targets for hackers. Patients' expectations to receive online services have forced healthcare systems to use both network-based services and cloud technology to provide easy access and instant availability of healthcare data. For this analysis we used the hypothetical iTrust healthcare database application to identify security risks for EMR data.Our analysis shows that the introduction of the Emergency Responder (ER) requirement poses the highest security threat to the iTrust database application. The CIA triad revealed the two main concerns for the iTrust database application: integrity and confidentiality.

Keywords: iTrust, risk analysis, ease of attack, vulnerability, threat, integrity, healthcare data

Introduction

With increasing cost of care, healthcare providers are looking for more cost-effective ways to provide patient services, such as developing EMR systems. Advances in technology allow healthcare providers, payers, and patients access to healthcare data at any time and from anywhere using the Internet[1]. Using this approach, providers can create their own EMR databases and share them with other EMR systems providers to make a distributed data center, the so-called electronic health records (EHR) system, to cut costs and improve healthcare big data. EHRs provide easy and instant access to healthcare data that can aid in patients' treatments and manage diseases on micro and macro levels, among other benefits. Despite all the positive effects, the major concerns of highly accessible healthcare data are privacy and security.

1.0. Situational Analysis

This paper explores issues about privacy and security of healthcare data based on the hypothetical iTrust healthcare application. After a thorough review of the data present in iTrust's database tables, we calibrated the value of each table and placed them into one of three categories: low, medium, and high. The categories represent the value the data in the individual database has for a potential attacker and the impact of a potential data breach for the iTrust database application (Table 1).

1.1. Database Table Value Allocation

To fully understand the security threats introduced by the new iTrust requirements, we assigned value points to each database table that describe the importance of the data in each database table, both for the iTrust database application and for an attacker. Tables with scores closer to 100 are more valuable to iTrust and attackers than tables with value scores closer to one. For example, the *users* table has a team-agreed value of 100. This table contains usernames and passwords for the iTrust system, which is important to the daily operation of the iTrust system and is valuable to attackers. Conversely, the *cpt-codes* table has a team-agreed value score of three, and contains standard codes for medical procedures. While this table is necessary for iTrust to operate correctly, it can easily be recreated if deleted or modified.

Table 1. iTrust data sets categories based on value point scores

Impact	Database Table	Definition/Importance
Low Impact Category	▪ Cptcodes ▪ Hospitals ▪ ICD Codes ▪ Lab procedure ▪ ND codes	This category contains mostly standard publicly available information in a key-value pair format that could easily be replaced and recreated if modified or destroyed. The exception to this general categorization is the *lab procedure* table, which contains a collection of IDs and codes referencing completed lab procedures.
Medium Impact Category	▪ Allergies ▪ Office medication ▪ Log in failures ▪ Office procedure ▪ Office visits ▪ Office survey ▪ Office diagnosis ▪ Transaction log	Data in this category is typically more complex than simple key-value pairs.
High Impact Category	▪ Patients ▪ Personal health information ▪ Personnel ▪ Users	Unauthorized access to these tables would disclose confidential information. For example, the *users* database table contains the medical identification (MID) and associated passwords of each user of the iTrust software.

Table 2. Individual and team value-point for iTrust datasets

No.	Database Table	VG	NK	IP	HZ	Mean	Adjusted by the Team
1	Allergies	3	40	3	40	21.5	20
2	Cptcodes	1	5	2	8	4	3
3	Hospitals	1	1	5	8	3.75	5
4	ICD codes	5	5	2	3	3.75	5
5	Lab procedure	5	13	8	8	8.5	8
6	Log in failures	20	1	13	2	9	13
7	Nd codes	1	3	2	2	2	2
8	Office visits	20	8	8	13	12.3	13
9	Ovdiagnosis	13	13	13	20	14.8	13
10	Ovmedication	20	2	5	40	16.8	20
11	Ovprocedure	20	2	3	20	11.3	13
12	Ovsurvey	13	20	1	20	13.5	13
13	Patients	100	40	100	40	70	100
14	Personal health information	100	100	20	100	80	100
15	Personnel	100	40	100	40	70	100
16	Transactionlog	40	1	13	20	18.5	20
17	Users	100	100	100	100	100	100

1.2.1. Computing value points*:* To reach these values, each author reviewed the content of the tables and made a determina-tion on the tables' values. After each author had a compiled list of values for each table, the group averaged and discussed the final scores. Value points were limited to the discrete choices of 1, 2, 3, 5, 8, 13, 20, 40 and 100, with duplicates allowed.

Because averaging scores resulted in values outside the allowed value points, the team discussed whether to round values up or down. Through this discussion and averaging, the team-agreed values were compiled, as seen in the "Agreed Team Value" column in Table 2. These values were then used to determine the security risk introduced by each requirement.

2. Introduction of Requirements with "Ease of Attack" Points

2.1. Emergency Responder

According to the iTrust Case Study, one of the new requirements to be implemented is the addition of a new role: Emergency Responder (ER). The role's scope is defined as "police, fire, emergency medical technicians (EMTs), and other medically trained emergency responders who provide care while at, or in transport from, the site of an emergency."[2] The ER will have access only to information important in emergency situations.

The information an ER will have access to include "allergies, blood type, recent short-term diagnoses, long-term, chronic illness diagnoses, prescription history, and immunization history."[2]. The introduction of the new role of ER into the existing program environment will likely increase existing security challenges for the iTrust database., for example:

- The ER role will need access to data stored in 10 different data tables, three of those databases (*patient, personal health information,* and *personnel*) have value points of 100, identifying them as high value databases with confidential information that is highly interesting for a malicious attacker.
- The role of an ER will give access to the iTrust database to a wide variety of different ER personnel, working for public as well as private health organizations. While all relevant ER providers are almost certainly covered under the HIPAA privacy rule and eligible to have access to PHI, access can be given only on a need-to-know basis.[3]
- Personnel working for those ER providers will likely heavily fluctuate. User accounts and privileges for the ER role have to be maintained and updated accordingly.
- ER will need access to the data from a wide variety of stationary and mobile computer systems, including mobile devices.

The access needed to enable ER to have all the information they need in an emergency will require special security measures to protect confidential data, specifically PHI and PII. Threats associated with the new ER role include, but are not limited to:

- Granting of excessive privileges/privilege abuse by ER personnel
- Malware or other malicious software (on devices used by ER personnel to access the iTrust database),
- Sniffing and man-in-the-middle attacks on unsecure internet (connections used to access the iTrust database)

- The disclosure of user IDs and passwords through negligence and misconduct
- Loss or theft of devices with confidential information
- Malicious insider attacks by ER personnel. Malicious insiders are one of the biggest threats in cybersecurity and, according to 2014's "Data Breach Investigation's Report," account for 15% of known incidents[4].

Based on the high vulnerability of ER requirements we allocated 100 "ease of attack points" to the implementation of the ER role. The high amount of "ease of attack" points" in correlation with necessary access to high-value database tables means the ER role is the requirement ranked as the highest security risk of all new requirements. To mitigate the security risk related to the ER role, iTrust should consider:

- Applying mandatory multi-factor authorization for ER personnel
- Robust password requirements
- Recurring security and threat education emails to all ER personnel
- Mandatory background checks from ER providers for their personnel
- Frequent updates to the ER personnel's access privileges to the iTrust databases. Furthermore, all information, including usernames and passwords, should be encrypted to prevent sniffing and man-in-the-middle attacks on unsecure devices and Internet connections.

To ensure that only ER personnel have access to the data they need in emergency situations, access must be given based on the principle of least privilege. One way to combat this issue would be to create a separate database. Such a database would include the most basic of patient information: allergies, blood type, recent short-term diagnosis, long-term/chronic illness diagnosis, prescription history, and immunization history[2].

Sensitive patient information should not be included in this viewable information. Patients can be identified by legal name and their birthday. Additionally, patients should retain the right to know that their data was accessed. This new database should be designed with security in mind. "Security should be foreseen as part of the system from the very beginning, not added as a layer at the end"[5].

2.2. Find Qualified Licensed Health Care Professionals

The second new requirement we had to implement was enabling patients to find qualified licensed healthcare professionals (LHCP) based on their expertise and location. According to the iTrust case study[2], patients will have access to the following LHCP information: name, contact information, number of patients treated for that diagnoses, list of all prescriptions given for that diagnosis, list of all laboratory procedures ordered for that diagnosis, visit satisfaction, and treatment satisfaction[2].

The data tables identified as necessary to ensure the availability of the demanded information are: *labprocedure, officevisits, ovdiagnosis, ovmedication, ovprocedure, ovsurvey,* and *personnel*. The *personnel* data table has 100 value points identifying the high value of the *personnel* data for a possible attacker. We allocated 100 "ease of attack points" to this new requirement. Because of the high value of *personnel* data and necessary access to data from seven different data tables, this requirement reaches the same security risk rating as the ER requirement; however, as it will need access to less high-value databases compared to the ER requirement, this requirement is ranked the second highest security risk.

The main security risks for this new requirement— comparable to the implementation of the ER role—are the vast number of possible devices and possible unsecure Internet connections used to access data from a high-value data table such as *personnel*. Possible threats include malware or other malicious software on devices used by patients to access the iTrust database, sniffing and man-in-the-middle attacks on unsecure Internet connections, and the disclosure of user IDs and passwords through negligence and misconduct, and loss or theft of devices with user information.

2.3. Update Diagnosis Code Table

The third new requirement is preparing the iTrust database for the introduction of new codes for diagnoses. According to the iTrust, the American Medical Association updated codes from *ICD-9CM* to *ICD-10-CM* since 2015. The addition of the new *ICD-10-CM* code to the iTrust database requires written access to the *icdcodes* data table. The data in the *icdcodes* table is not confidential and not of high interest for a malicious attacker. Trusted personnel, will almost certainly be responsible for adding the new codes. This limits the threat of a malicious attack or misconduct in the process. We allocated three "ease of attack" points to this requirement. Although the *icdcodes* table has 40 value points in combination with the low "ease of attack" points, this requirement is ranked as the least security risk of the new requirements (Table 3).

Table 3. iTrust security risk analysis

Requirements	Ease of Attack Points	Databases	Max Value Points	Security Risk	Rank of Security Risk
Emergency Responder	100	1,2,4,7,9,10, 11,13,14,15	100	10000	1
Find qualified LHCP	100	5,8,9,10, 11,12,15	100	10000	1
Update code table	3	4	13	39	3
View access log	40	13,15,16	100	4000	2

Table 4. CIA Risk Assessment Analysis

	Loss of integrity	**Loss of availability**	**Loss of confidentiality**
Summary of score for all datasets	485	330	360
Maximum Possibility	1700	1700	1700
Losing Probability	28.5 %	19.4%	21.2%
Vulnerability Rank	1	3	2

However, the update of the diagnoses codes still constitutes a security risk to the iTrust database application. According to TechTarget (2014), ICD-9CM contains ~13,000 three- to five-digit diagnosis codes, while ICD-10CM has more than 68,000 seven-digit codes[6]. Challenges implementing a change like this could likely cause data links and processes to fail, creating possible security threats. The iTrust administrator will have to identify all systems and processes; the change will affect and test transactions involving the new ICD-10CM codes. If an outsourced software provider does the update, timelines for the upgrade must be identified to not impede the functions of the database. In any case, a database backup should be done before the update.

2.4. View Access Log

The last new requirement is the addition of a view access log, enabling patients to view "the names of licensed healthcare professionals that viewed or edited their medical records and the date the viewing/editing occurred is displayed"[2]. To provide the requested information, the view access log requirement needs access to data in the *personnel, patients*, and *transactionlog* data tables. The main security risk for this new requirement is similar to the "find qualified licensed health care professional" requirement as the risks are based on the same user group.

3. Additional Considerations

Online accessible databases are known to be vulnerable to SQL injection attacks. iTrust has a known record of issues with SQL injection attacks, which stem from insecure codes and "improper neutralization of special elements used in a SQL command"[7, 8]. According to the iTrust Compliance Report, SQL injection attacks top iTrust's list of software errors and vulnerabilities[9].

Additionally, allowing users to make their own changes to the database can be risky in itself, so Access Controls are another method of controlling who exactly is allowed to edit the database, and what kind of changes they are permitted to make[10].

Robust password requirements will be an important tool to protect the iTrust database. To improve the protection, multi-factor authentication should be implemented. A proper authorization process in combination with strong authentication allows the administrator to configure user accounts' access permissions. Authorization controls what iTrust data users can or cannot access[11].

For an online accessible database like iTrust, encryption is especially important. Not only PHI, PII and other confidential user data are threatened, user names and passwords could also be revealed if submitted without encryption[12]. HIPAA/HITECH regulations define the encryption needs to meet FIPS-140-2 requirements to protect iTrust from fines if data are stolen[8, 13, 14]. The encryption needs to address mobile security additional to standard disk and file encryption.

4. Discussion: Health Information Is Valuable

"(About) 47% of the U.S. population have had their personal healthcare data compromised over the last 12 months"[15]. All health information is extremely valuable to hackers[16]; an individual's stolen medical records are usually sold or used to obtain drugs, treatment, and/or medical equipment, usually at the victim's expense[16]. Victims usually find out about the theft months or even years after it has taken place, generally when they receive a bill for items or services they never authorized[16]. Credit cards are typically canceled immediately after a compromise is suspected, but medical and personal data is much more difficult to change [17]. The Medical Identity Fraud Alliance's 2015 study showed that victims of medical identity fraud pay an average of $13,500; health records are heavily regulated and correcting information that has been tampered with is extremely time-consuming and expensive[16]. Another angle hacker can take with stolen medical data is, to seek a ransom from the medical institution the data was originally stolen from. Medical institutions aim to

preserve their reputation, which means not only failing to report security breaches but to also attempt to cover them up (e.g. 2016, medical centers in Hollywood, California and Australia) [16]. Hackers also use medical data to commit tax fraud[16].
State-sponsored espionage is also a possibility when it comes to stolen medical information. Anthem, experienced an attack that resulted in the theft of 78 million customers' medical records[18]. Stolen information can be combined from multiple sources to create dossiers of both individuals and target groups[18].

5. CIA Triad as an Alternative Approach to Assess Threats and Vulnerabilities

Table 4 shows the likelihood of a potential successful attack on a company is measured in regards to confidentiality, integrity and availability [19, 20]. We calculated the potential threats, the confidentiality, integrity and *availability* concerns of all 17 databases. The scores for each CIA triad component were then summed to create a maximum possibility score of attack, shown in the Maximum Possible row in Table 4. The CIA triad revealed the two main concerns for the iTrust database application: integrity and confidentiality.

6. Conclusion

The process of adding requirements to the iTrust database application is nontrivial. The potential impact of the new requirements needs to be carefully evaluated and implemented with all stakeholders and actors in mind (including malicious ones). Security policies and controls will need to be updated and refreshed as new requirements come into play. After evaluation, the requirements themselves may also need to be updated or modified to better align with the security needs of the iTrust database application. The confidentiality, integrity and availability of the system should remain constant through the implementation of the new requirements. If changes to any of the application's security standards are expected, prioritized mitigation plans should be created with the aim of minimizing negative impacts. It is only when the potential impacts are understood that an organization such as iTrust should move forward with implementing new requirements.

References

1. Gruman G, Knorr E: **What cloud computing really means.** *InfoWorld* 2008, **37:**13.
2. UMUC: **iTrust Case Study. CSEC 630 Team Project. Archived at https://learn.umuc.edu/d2l/le/content/143544/viewContent/6440867/View.** 2016.
3. HHS.gov: **HIPAA for professionals. Health Information Privacy. Retrieved from http://www.hhs.gov/hipaa/for-professionals/index.html** 1996.
4. Diana A: **10 ways to strengthen healthcare security. Information Week. Retrieved from http://www.informationweek.com/healthcare/security-and-privacy/10-ways-to-strengthen-healthcare-security/d/d-id/1306631** 2014.
5. CERN: **Security checklist for software developers. CERN Computer Security Team. Retrieved from https://security.web.cern.ch /security/recommendations/en/checklistfor_coders.shtml** 2016.
6. **FAQ: How will the transition to ICD-10 code affect health IT? Retrieved from http://searchhealthit.techtarget.com/tutorial/FAQ-How-will-the-transition-to-ICD-10-codes-affect-health-IT#**
7. Zare H, Azadi M, Olsen P: **Techniques for Detecting and Preventing Denial of Service Attacks (a Systematic Review Approach).** In *Information Technology-New Generations.* Springer; 2018: 151-157
8. Fernandez EB, Alder E, Bagley R, Paghdar S: **A Misuse Pattern for Retrieving Data from a Database Using SQL Injection.** In *BioMedical Computing (BioMedCom), 2012 ASE/IEEE International Conference on*. IEEE; 2012: 127-131.
9. Wees A, Rogalski B, Zhang K, Scaramuzzino S, Root T: **iTrust Database Software Security Assessment. iTrust Database Software Security Assessment. Security Champions Corporation (fictitious) Assessment for client Urgent Care Clinic (fictitious). University of Maryland University College. Retrieved from: https://researchedsolution.wordpress.com/2013/09/14/itrust-database-software-security-assessment/.** 2013.
10. Goodrich M, Tamassia R: *Introduction to computer security.* Addison-Wesley Publishing Company; 2010.
11. Microsoft: **Security authorization. Retrieved from https://www.iis.net/configreference/system.webserver/security/authorization** 2016.
12. Oracle: **Oracle Eloqua Data Privacy Security Add-on Cloud Service, Configuration Guide. Retrieved from: https://docs.oracle.com/cloud/latest/marketingcs_gs/OMCAA/pdf/OracleEloqua_DataPrivacy_ConfigurationGuide.pdf.** 2017.
13. **HIPPA compliance for data encryption. Information security. Retrieved from https://frsecure.com/blog/hipaa-compliance-for-data-encryption/**
14. **Why encryption is crucial to your organization. Heathcare IT News. Retrieved from http://www.healthcareitnews.com/why-encryption-crucial-your-organization**
15. **Hacking healthcare IT in 2016: Lessons the healthcare industry can learn from the OPM breach. Institute for Critical Infrastructure Technology. Retrieved from http://icitech.org/wp-content/uploads/2016/01/ICIT-Brief-Hacking-Healthcare-IT-in-2016.pdf**
16. Dimov D, Juzenaite R: **Healthcare hacking. InfoSec Institute. Retrieved from http://resources.infosecinstitute.com/healthcare-hacking/** 2016.
17. Humer C, Finkle J: **Your medical record is worth more to hackers than your credit card. Reuters. Retrieved from http://www.reuters.com/article/us-cybersecurity-hospitals-idUSKCN0HJ21I20140924** 2014.
18. **Why hackers want your health care data most of all. InfoWorld. Retrieved from http://www.infoworld.com/article/2983634/security/why-hackers-want-your-health-care-data-breaches-most-of-all.html**
19. Swanson M: **Security self-assessment guide for information technology systems.** DTIC Document; 2001.
20. Stoneburner G, Goguen AY, Feringa A: **Sp 800-30. risk management guide for information technology systems.** 2002.

Alaska Department of Health and Social Services Cybersecurity Recommendations

Jennifer J. Stubblefield
University of Maryland University College, Adelphia, MD, USA
Jstubblefield2@student.umuc.edu

Summary— The Alaska Department of Health and Social Services (DHSS) current cybersecurity program needs updating due to recent cybersecurity breaches. The recent data breach is the second in less than ten years that resulted in over 500 patients' protected health information (PHI) being released. Currently, DHSS has several obsolete computers and programs, four divisions handle Medicaid claims, password criteria is weak, and there are outdated procedures and policies. IT needs to develop a replacement program for current hardware and software, which does not allow secure transmissions or security updates. If eCommerce is streamlined and Medicaid claims moved to a central office, then IT could prioritize that department in the upgrade. A password-cracking lab demonstrated that the state password criteria should be strengthened and the purchase of Ophcrack password cracker is recommended. Implementing these additional administrative, technical, and physical safeguards will increase the efficiency of the department while securing vital PHI.

Keywords— Password cracking; protected health information; eCommerce; cybersecurity

I. INTRODUCTION

The Alaska Department of Health and Social Services (DHSS) must increase its cybersecurity protocols to prevent hacking. DHSS had the second most significant settlement over protected health information (PHI) being released in 2009 and had once again released PHI. The governor has recognized a need for change by restructuring the state information technology (IT) office, which is still in process. The department must prioritize its IT office's needs to prevent additional violations.

Malware entered the DHSS computer system in two separate July 2017 incidents which compromised over 500 individuals' PHI [2]. The compromise occurred when two Office of Children Services (OCS) employees opened emails containing malware [2]. IT quickly mitigated the attack by isolating the computers, but PHI had been compromised [2]. The Medicaid division had a security breach in 2009, in which a USB drive was stolen from a DHSS computer technician's vehicle [5]. The drive contained over 2,000 patients' electronic protected health information (ePHI) and ended in a $1.7 million Health Insurance Portability and Accountability Act (HIPAA) settlement with the United States Department of Health and Human Services [5]. To prevent similar attacks, DHSS needs to upgrade computer systems and protocols.

II. INFORMATION SYSTEM INFRASTRUCTURE

Alaska Governor Bill Walker signed an administrative order establishing the Office of Information Technology (OIT) and creating the first state Chief Information Officer (CIO) in April 2017 [7]. IT staff is currently divided among 15 departments, but OIT plans to eliminate duplicate work, establish standard policies, and oversee all IT equipment [9].

DHSS is organized into four main divisions, which all operate under the same IT guidelines. Due to dealing with PHI, the divisions also abide by the Alaska Health Information Technology Plan [1].

The divisions are public health; Medicaid and health care policy; family, community, and integrated services; and, finance and management services.

Located in the Division of Finance and Management Services, DHSS OIT has five offices: security, business applications, customer service, network service, and strategic planning (Figure 1). In 2012, the office had 128 information technologists working with 4,000 Central Processing Units (CPU) for 3,600 employees working in 35 communities with 128 facilities [8]. The department currently still reports to the Finance and Management Services assistant commissioner, but must also abide by any policies established by the newly created state administration IT office.

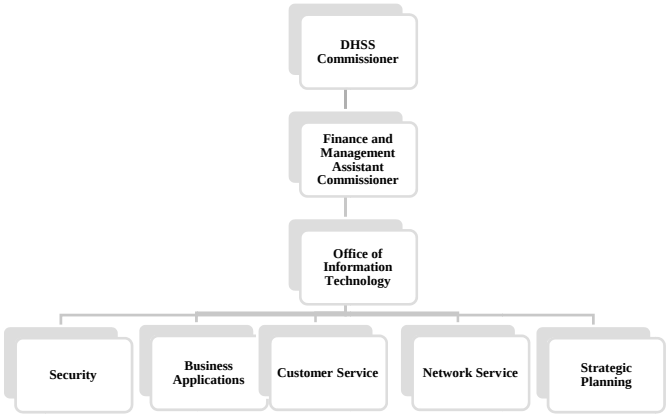

Figure 1. Alaska Department of Health and Social Services Office of Information Technology Organizational Chart. Information for chart retrieved from http://dhss.alaska.gov/fms/its/Pages/Default.aspx.

Currently, Alaska has both 32-bit computers and 64-bit computers with a Windows XP operating system (OS). A 64-bit CPU can support additional RAM and multiple cores allowing the CPU to perform faster calculations for more advanced graphics and video programs [3]. Software on all computers includes Microsoft Office 2010 and Norton utilities. Norton provides a network firewall, scans computers for viruses, and alerts users of dangerous websites, suspicious files, malware, and other cyber threats. Data storage and backup data are on local office servers. Each computer has a data recovery tool. Alaska contracts through JP Morgan to use Quest electronic debit cards to provide families with needed assistance [4]. The state has implemented a health information exchange network with public, private, and government offices that require additional security protocols (Figure 2). The system is vast, but essential items that need addressing include hardware and software.

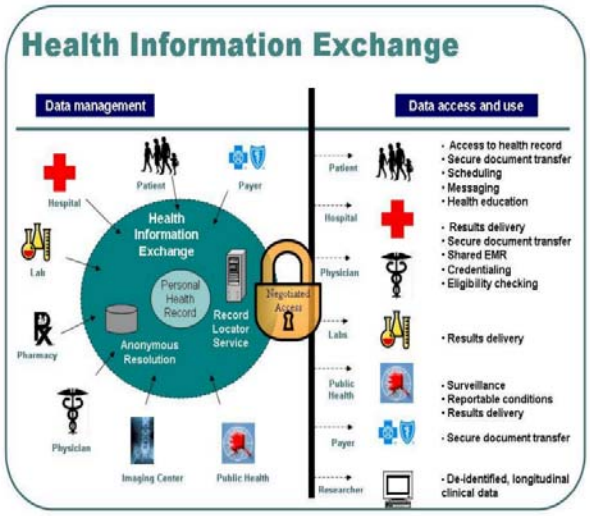

Figure 2. Alaska Health Information Exchange plan [1].

DHSS operates on a wide area network as it spans across the state (Figure 3). Alaska uses the hybrid network typology, which is a hybrid of the star and tree typologies [6]. Plus, the wireless typology is used to access the network from remote locations. These hybrid typologies prevent the entire network from being shut down due to an isolated server problem.

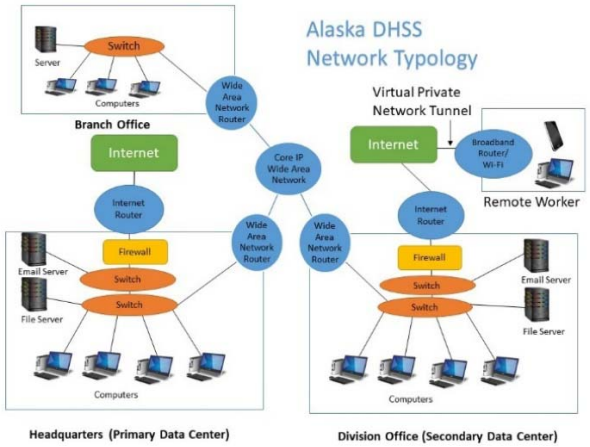

Figure 3. Alaska Department of Health and Social Services Wide Area Network.

The department protects the confidentiality, integrity, and availability of PHI by multiple methods. First, the division implemented cyber policies and procedures to protect information by addressing the storing, accessing, and transferring of information. Secondly, the network has a computer security software firewall preventing viruses from being downloaded and hackers to gain access. Data is stored encrypted on servers in locked rooms or cabinets with building security. There are primary and secondary data centers established. These administrative, technical, and physical safeguards implemented provide protection.

III. IDENTITY MANAGEMENT

Identity management (IM) includes user authorization, authentication, and access control management. User access is determined by authorization through department managers. Any person

using the state website login or computers are assigned a username and temporary password. When logging in, the computer will authenticate the user. Protect eCommerce programs by requiring multi-factor authentication, such as requiring a second password. Next, DHSS has an access control list, and only those in the department have access to shareable files. Additionally, there are role-based access controls. For example, supervisors have access to their employees' personnel files. In file access control, the owner of the file will determine who has access. These owners review the access list quarterly. Database access control includes, for example, an employee assigned to children services division will have access to the foster facility database. Mobile access control consists of access through a designated sign-on from a remote laptop, tablet, smartphone, or another mobile device. Mobile users are only allowed access to specific databases. User authorization, authentication, and access control management are vital to protecting PHI.

Password management is an integral part of cybersecurity. Currently, state employees are required to have a password with a length of 8 characters including at least two of the following, lowercase characters, uppercase characters, and numbers. Employees have noted that often the same password is used, and the number changed at the end consecutively when resetting (i.e., Motorcycle1, Motorcycle2, etc.). A simple password change can be easily identified by password cracking programs. Therefore, the state should allow a password cracking tool to be used by IT and implement stronger password criteria.

After lab testing both Cain and Abel and Ophcrack password programs, Ophcrack was the most straightforward program to use, but Cain and Abel cracked more. However, the longest lab passwords were ten characters, and most were four. The Cain dictionary attack broke two 10-character passwords using only letters within seconds. The Cain brute force attack took the most time and identified the estimated time left to crack the password. One password that could have up to 16 characters had 8.69741e+010 years left to crack, and another password with up to 8 characters and numbers had 297.757 days left. Cain and Abel also required each user to be selected one at a time. Ophcrack allowed selecting all users simultaneously and within seconds returned results. Both cracking tools recovered only 23% of the passwords, which is a total of 12 of 53. Brute force attack of up to eight characters, including upper and lowercase letters and numbers, cracked five passwords of 13 users. This lab demonstrates the need for a stronger password criterion.

The password cracking tools have benefits and risks. Benefits include detecting cryptanalysis attacks, certificate spoofing, and password recovery, but the programs are identified as malware by security programs. On Wi-Fi networks, the same cracking tool can be used by hackers to capture passwords. To prevent clear text passwords, IT can use secure socket layers (SSL), transport layer security (TLS), and similar security features. The benefits of the additional protection outweigh the risks of cracking tools. Ophcrack is recommended for the state network as it is compatible with the current operating system and is more efficient.

IV. HANDLING RISKS

First, some risks must be accepted. Due to budget constraints, DHSS has not kept up with technology. Some of the clinics lack any technology, and public health nurses have to be trained on paper charts [8]. The state should use mobile tablets or e-readers for health providers working remotely. Also, the state is reassessing its wide array of computers. IT should stay informed and develop a budget plan. If risks are allowed with no action, the system will continue to be compromised. Therefore, accept risks such as budget constraints, but plan for long-term change through budget proposals.

Second, transfer risks. Four offices handle Medicaid claims, and outdated equipment does not allow cross-referencing. Streamline work by processing Medicaid claims in one office. Also, hiring

cybersecurity firms in securing data or analyzing existing platforms allows state workers to concentrate on current projects. Transfer risks by hiring contractors and streamlining offices handling PHI.

Third, IT can mitigate risks. A security program can be configured to find anomalies identifying an individual's suspicious behavior, restrict user access to specific files or folders, require stronger passwords, and disable the ability to upload executable programs. Employees must be trained in cyber threats. Policies and procedures should include new technology. Strengthening security controls, password access, and policies can mitigate cyber threats.

Finally, eliminate risks. IT deals with hardware and software that is no longer supported by manufacturers. The DHSS travel budget is $4 million annually [8]. If IT upgrades video conferencing allowing the travel budget to be cut 10% initially, this will give the department $400,000 for the computer upgrade project. A recent quote estimates $200,000 to immediately replace 87 obsolete computers and upgrade software. Eliminating outdated equipment will remove security risks.

V. CONCLUSION

DHSS had a security breach that identified weaknesses in protecting PHI. The recent data breach is the second in less than ten years. DHSS has worked on significant plans to prevent PHI violations. IT staff must begin executing plans including replacing outdated computer and software, streamlining eCommerce into central offices, establishing current cyber user procedures and policies, strengthening existing security measures, and training employees in cyber threats.

References

[1] Alaska Department of Health and Social Services. (2010). Alaska health information technology operations plan. Retrieved from http://dhss.alaska.gov/hit/documents/hit_operations_plan.pdf

[2] Alaska Department of Health and Social Services. (2017, September 1). HIPPA breach notification. Retrieved from http://www.dhss.alaska.gov/News/Documents/press/2017/HIPAA%20Breach%20Press%20Release%20WEB.pdf

[3] Computer Hope. (2017, April 26). What is the difference between a 34-bit and 64-bit CPU? Retrieved from Computer Hope: https://www.computerhope.com/issues/ch001498.htm

[4] Division of Public Assistance. (2015). Alaska quest cards. Retrieved from DHSS: http://dhss.alaska.gov/dpa/pages/ebt/default.aspx

[5] McCann, E. (2012, June 27). Healthcare IT news. Retrieved from http://www.healthcareitnews.com/news/alaska-pays-17m-hhs-data-breach

[6] Mitchell, B. (2017, August 21). Computer network typology, illustrated. Retrieved from Lifewire: https://www.lifewire.com/computer-network-topology-illustrated-4064043

[7] Office of the Governor. (2017, April 25). Governor Walker announces overhaul of state IT services. Retrieved from https://gov.alaska.gov/newsroom/2017/04/governor-walker-announces-overhaul-of-state-it-services/

[8] Public Works LLC. (2015, September 28). Alaska Department of Health and Social Services performance review. Retrieved from http://www.akleg.gov/basis/get_documents.asp?session=29&docid=51398

[9] Walker, B. (2017, April 25). Administrative order no. 284. Retrieved from Governor Bill Walker: http://doa.alaska.gov/oit/docs/AO-284-Establishing-Office-of-Information-Technology.pdf