Secure Research Infrastructure Using tiCrypt

Erik Deumens
deumens@ufl.edu
University of Florida
Gainesville, Florida, USA

Thomas Samant
tsamant@ufl.edu
University of Florida
Gainesville, Florida, USA

Eric Byrd
ebyrd@umd.edu
University of Maryland
College Park, MD, USA

Samuel Porter
sporter1@umd.edu
University of Maryland
College Park, MD, USA

Axel Haenssen
axel@princeton.edu
Princeton University
Princeton, New Jersey, USA

Curtis W. Hillegas
curt@princeton.edu
Princeton University
Princeton, New Jersey, USA

Elizabeth Holtz
eparham@princeton.edu
Princeton University
Princeton, New Jersey, USA

Irene Kopaliani
ik8@princeton.edu
Princeton University
Princeton, New Jersey, USA

ABSTRACT

Digital data is increasingly important and becoming more prevalent in research. As such, a significant and growing fraction of that data is subject to complex security and privacy standards for its storage and processing. With the introduction of contractual requirements for NIST Special Publication 800-171 and other compliance frameworks, universities are under expanded pressure to secure research infrastructure to ensure their researchers comply with laws, regulations, and contractual requirements.

The University of Florida developed tiCrypt, a bundled security middleware product, and implemented their ResVault system to meet the growing need for secure research infrastructure. The University of Maryland and Princeton University, both seeking to meet compliance requirements for secure research, have also implemented systems using tiCrypt. As new institutions adopt tiCrypt, the Tera Insights development team continues to add functionality to meet the needs of a broad range of research workflows.

CCS CONCEPTS

• **Security and privacy** → **Public key encryption; Security requirements; • General and reference** → *Design*; • **Computer systems organization** → Special purpose systems.

KEYWORDS

CUI, NIST SP 800-171, FISMA, NIST SP 800-53, HIPAA, DUA, DFARS, VDI, secure research infrastructure, tiCrypt

ACM Reference Format:

Erik Deumens, Thomas Samant, Eric Byrd, Samuel Porter, Axel Haenssen, Curtis W. Hillegas, Elizabeth Holtz, and Irene Kopaliani. 2021. Secure Research Infrastructure Using tiCrypt. In *Practice and Experience in Advanced*

This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

PEARC '21, July 18–22, 2021, Boston, MA, USA
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8292-2/21/07.
<https://doi.org/10.1145/3437359.3465577>

Research Computing (PEARC '21), July 18–22, 2021, Boston, MA, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3437359.3465577>

1 INTRODUCTION

The use of restricted data in research proliferates both in the number of projects using such data and in the scale of the data sets. This data includes Personally Identifiable Information (PII), Protected Health Information (PHI), data covered by HIPAA[1], FISMA[4] or Data Use Agreements (DUA), data with Controlled Unclassified Information (CUI) requirements, and data that is the intellectual property (IP) of the PIs and the institution. Traditionally, when requests were made to host restricted data and analysis on central research computing and data resources, a one-off solution would be designed and implemented for each project. As the University of Florida (UF), the University of Maryland (UMD), and Princeton University considered this approach, they concluded that such a strategy was not scalable nor did it provide a responsive experience for the researchers.

By ruling out the traditional approach to managing the use of restricted data in research, UF, UMD and Princeton had to define the scope of a successful strategy. All three institutions were seeking a scalable solution for secure research infrastructure. The desired system needed to scale from small to large datasets and from modest to demanding computational requirements. The system should be purpose built for providing security around research data and computation plus it should fit into and complement existing computational and data infrastructure. It should be manageable by a small team. The system needed to be auditable and ready for assertion or certification of compliance by a third party to satisfy regulation and compliance requirements. Most importantly, the universities needed a tool that would be accessible to and easy to use by researchers, providing a positive and productive user experience. In addition to meeting these requirements, it was also expected that this service would be provided at a manageable cost by a trustworthy vendor.

The tipping point for all three universities came in the summer of 2015 when the federal government published a draft of a new Federal Acquisition Rule (FAR) requiring that all federal data

classified as CUI must be treated by non-federal organizations as described in NIST Special Publication 800-171. The Department of Defense (DoD) immediately declared that the draft requirements would be in the updated version of the Defense Federal Acquisition Rule Supplement (DFARS). After strong push back from the community, DoD agreed that the new rule would become effective in December 2017. This brought the term CUI to the academic world's attention. It was first apparent in the Department of Defense and export control arena and then, because of the pending FAR (which is still pending in 2021), to all other activities leveraging federal data. For example, the Department of Education has issued a memo that recommends treating student data as CUI. This was recently reiterated with regards to Federal Student Aid (FSA) data.[2]

This paper will introduce tiCrypt, a product developed at the University of Florida and provided by Tera Insights, as the selected tool that met all requirements for handling restricted data, and it will detail how each of the involved universities has adopted and implemented tiCrypt to provide a secure research infrastructure for researchers and their collaborators. Distributed throughout the paper, Tables 1, 2, 3, 4, and 6 are use cases from each institution. While all three institutions had similar motivators that led to the implementation of tiCrypt, the technical aspects of each implementation vary. This speaks to the flexibility of tiCrypt to meet the unique needs through architecture design.

2 TECHNICAL DESCRIPTION OF TICRYPT

tiCrypt uses RSA-2048 keys with AES-256 symmetric algorithms for encryption to protect resources and secure communication channels. Due to the user-controlled end-to-end encryption, system administrators do not have access to user's data unless explicitly shared by the end-user.

Users access deployed versions of the software via an application called "tiCrypt Connect." This application serves the front-end locally in a browser and works in tandem with the front-end to establish secure VM connections (See Fig. 1). [12][13]

tiCrypt users generate a public-private key pair upon registration which is protected by a user-created strong password. All the security mechanisms in tiCrypt use the public or private key to perform cryptographic encryption, decryption, and key exchanges.

2.1 End-to-end encryption and zero-trust

When a user creates an account within tiCrypt, a public-private key pair is generated in the browser. The key pair generated is derived from a user-set account password. The public key is stored on the server, while the private key is in the user's sole possession. The private key is always needed for a user to log into the system. The only time a password is used is to decrypt the private key.

Additionally, public-key cryptography is utilized for secure data storage and sharing. Every file uploaded into tiCrypt is encrypted with a unique AES-256 key. Those AES-256 keys are then encrypted with the public key of the user or groups of users.

File sharing occurs via secure key exchange in the browser of the file owner as opposed to a server operation:

- (1) The encrypted key is accessed from the server.

Table 1: Case Study - ITAR/EAR engineering workflow

A team in the Herbert Wertheim College of Engineering at UF is working with a company on the design of a circuit board that is advanced technology under export control. The project team does the design using CAD software in a ResVault VM. When the design is satisfactory for testing, it is transferred in a secure manner to the company that will manufacture the board and ship it to UF. The board is then installed in a dedicated computer in an ITAR compliant lab that is connected by UF's CUI compliant virtual network to ResVault. The board is tested and any data collected is sent to ResVault over the secure network for analysis and to inform the next iteration of the board's design.

- (2) The encrypted key is decrypted using the owner's private key. Then the key is encrypted with the destination's public key.
- (3) The key encrypted with the destination's public key is deposited on the server.

2.2 tiCrypt user components

tiCrypt has two main end-user service components. The first component, called the Vault, is for cold storage like workflows and ad-hoc file sharing.

The second component is virtual machines providing virtual desktop infrastructure processing and analysis-oriented workflows. Similar security mechanisms based on public-private key pairs apply when users access a virtual machine in the tiCrypt environment. For example, the VMs access the central data store through sub-sessions established by the client, with the key exchange occurring through an encrypted user to VM channel. Additionally, all the communication between the browser running tiCrypt and the virtual machine occur using secure TLS based tunnels. [14][13]

3 DEVELOPMENT AND IMPLEMENTATION

3.1 UF's development of tiCrypt/ResVault

In 2015, the University of Florida (UF) signed a contract for a project that required FISMA compliance in order for the information system to store and host all processing of that data. The contract was significant to the university and the commitment was made to architect and build such a system. The goal was to meet the controls described in NIST Special Publication 800-53 at the moderate impact level. Various options, including using the commercial cloud, were explored. The decision was made to build a system using enterprise architecture with Virtual Desktop Infrastructure (VDI) to provide access and distinct security domains to keep various projects separate. The task met its deadline and, with massive effort, was completed in 90 days. Since 800-171 is a tailoring of 800-53 Moderate, the university was fortunate to have undertaken the effort, as it was now prepared for this forthcoming requirement. The University of Florida quickly learned that the enterprise architecture system implemented is labor intensive to operate for the multitude of smaller projects that require DFARS compliance. Therefore, in 2016, UF Information Technology embarked on a project to develop

Figure 1: tiCrypt Overview Created by Thomas Samant.

a system that would allow new projects to be more easily configured and provisioned. This project produced tiCrypt and built a system that is branded at UF as ResVault.[14]

The first principle of tiCrypt architecture is based on strong security through end-to-end encryption and zero-trust in all services. The second aligns with high-performance computing architecture where a powerful shared hardware platform provides storage and computer resources and different projects are allocated using a simple interface, rather than requiring custom builds by network and system engineers. The separation between projects is ensured by the encryption of all data and communication using public-private key infrastructure where only the user holds the private key. [12][13]

The initial design includes a workspace similar to Google Drive and Dropbox to allow data to be stored and shared securely. Next, the capability to allocate and run virtual machines (VM) for reading and writing virtual drives was added. This grew to the recent addition of features to configure virtual clusters inside the secure environment to allow projects to use standard design principles and software available in the open computing world. The VMs can run Windows or Linux and all communications are encrypted and strictly controlled. VMs can only talk to other VMs in the virtual cluster and cannot reach the outside world, unless explicitly granted access, like to check out a license. [14] [12][13]

tiCrypt is the security middleware of the system and was developed at UF by Prof. Alin Dobra. Tera Insights has the rights to sell the software. The first production version of a system running tiCrypt was reviewed by an external company for compliance with NIST 800-53 Moderate and 800-171 in 2018. [14]

3.1.1 UF's implementation of tiCrypt/HiPerGator-ResVault. The tiCrypt system has been operated by the Research Computing (RC) unit in UF Information Technology (UFIT) together with the Information Security Office (ISO) and the Infrastructure and Communications Technologies (ICT) units since 2017. The administrative management is done in close collaboration with the Office of Research and its Export Control department. A significant number of

Table 2: Case Study - Video review workflow

A team of students at UF need to review a collection of videos and create annotations. The content of the videos are PII and need to be protected accordingly. The videos are stored in the tiCrypt vault, part of the ResVault system at UF. A video viewer is available in the vault so that project participants, mostly students, can view the videos easily without needing to start a VM. The project manager can ensure that videos are shared only with the necessary project participants and no one else.

ITAR/EAR controlled projects have been migrated into ResVault. In addition, the system provides the main computing infrastructure for a number of projects working electronic protected health information (ePHI), IP, and PII.

The potential use of ResVault is triggered during contract review by the Office of Research or from a project risk review by the ISO. A call is then set up with the principal investigator and their partners and subcontractors, the research administrators, and the UFIT RC staff. Determination is made during this call as to whether the project workflows can be supported by the environment and what would need to be done outside of the environment. [14]

3.2 UMD's adoption of tiCrypt

University of Maryland began developing a strategy to handle data subject to DFARS 252.204-7012 in 2017. A secure and compliant computing environment was a strategic priority due to the significant proportion of research funding at UMD that can be attributed to the Department of Defense and associated prime contractors[16].

From the start, discussions revolved around some form of secure and virtualized environment, but it wasn't clear whether to host the environment on premises or in the cloud. In addition, there were few commercially available 'turn-key' systems that met the security requirements of NIST 800-171.

tiCrypt Constellation Architecture

Figure 2: Constellation Overview Created by Thomas Samant.

Table 3: Case Study - Multi-VM and instrument collaborative workflow
<p>A team at UMD tests materials for customers inside a specialized laboratory. The test involves several probes to gather measurements and several workstations dedicated to a specific function during a test. Multiple systems share access to files that have test run parameters, while another runs software to process results. Some VMs can only read processed data while others can write. tiCrypt enables this setup using its two unique features: Constellations and Forms. The Forms feature allows the probes to collect and submit a survey and form data directly into tiCrypt. With Constellations (See Fig. 2), a domain is configured to manage the group policy, user permissions, and expose the file system to all VMs in the group (through SAMBA). It enables shares between worker VMs which is not possible within the standard tiCrypt operation.</p>

Internal estimates and initial feedback from other institutions all indicated that hosting the environment in the cloud would be too costly. In addition, UMD had existing vendor relationships and extensive experience running on premises virtual infrastructure, but was still developing skill sets around cloud architecture. The effort moved forward to gather quotes and operational costs for creating a dedicated enclave on campus.

Between 2018 and 2019, internal staff turnover prevented much progress on the project, but the need remained critical. In 2019, after initial testing and piloting was done by IT staff in the School of Engineering, tiCrypt was selected to provide the layer of security necessary to create a compliant and secure computing enclave. Tera Insights agreed to work closely with UMD in order to leverage

Nutanix hyper-converged nodes as the underlying infrastructure. At the end of 2019, all hardware and software had been installed on campus in a secure physical location.

The factors that lead to the tiCrypt / Nutanix solution were:

- The hardware and software combine to create a layered approach to security
- tiCrypt is a zero-trust solution, which is key to good security architecture
- tiCrypt provides a high-quality user interface for the researchers
- As a whole, the solution could be configured and managed by a small team

UMD’s tiCrypt implementation runs on hyper converged hardware in a secured area of the on-campus data center. Hardware was installed by an implementation partner, and the network was configured in coordination with the UMD network engineering team. Tera Insights assisted with the initial installation of tiCrypt. Installation was followed by approximately five months of documentation and process design to develop a System Security Plan and standard operating procedures that would meet the NIST 800-171 requirements necessary to protect CUI. This work was done in partnership with an outside consultant. At the same time, the DIT Research Computing team began to form around one existing Senior Research Analyst and between December and March the Associate Director, Engineer, and Security and Compliance Analyst roles were filled.

From installation to campus launch took approximately six months. This time period included winter and spring break, the hiring of three team members, and a switch to remote work in response to the COVID-19 pandemic.

The tiCrypt system is supported by a team of four who are the primary resources for researchers with CUI and who need to meet DFAR 252.204-7012 obligations. The team works closely with UMD's Export Control Office and the Office of Research Administration to ensure researchers are aware of the available solution, to assist with contract questions related to compliance, and to develop Technology Controls Plans. Training is provided to all users prior to granting them access to the system. Given the goal of granting least functionality within the system, images are also tailored to the specific needs of researchers, often working with local IT and software companies to ensure the necessary software can be installed and activated.

The team worked closely with the ServiceNow team to add the tiCrypt system inventory and configuration into the ServiceNow Configuration Management Database that integrates with the Change Management process. This has helped ensure changes are appropriately documented and replaced the initial spreadsheet-based process.

It is anticipated that in the coming months additional system automation around building and deploying images as well as incorporating the system into the UMD IT services chargeback system will be introduced.

3.2.1 UMD's Technical Implementation. The University of Maryland's CUI environment is installed within a secured, on-campus data center, with multiple layers of physical access controls. A five-node Nutanix hyper converged virtual infrastructure is logically and physically separated from other campus computing infrastructure.

Network security is accomplished via host level firewalls, campus data center firewalls, and VLANs are utilized to separate network traffic. Access to the environment has been placed behind VPN groups.

The native Nutanix hypervisor solution, AHV, is used to build our virtualized environment. All the systems within CUIE are virtualized using AHV and managed with the Prism interface, a robust REST API. The application running on top of this hyper converged solution is Tera Insights tiCrypt. This is ideal because Nutanix AHV and tiCrypt both leverage the libvirt API and the KVM virtualization module.

The 'infrastructure' of the system consists of a Microsoft Deployment Toolkit (MDT) server to create images, Nexpose server for vulnerability scanning, Splunk forwarders, a Veeam AHV backup proxy, a Nvidia GPU license, and tiCrypt mailbox server.

Research VMs are based on hardened Windows 2019 and RedHat Enterprise 7 Operating Systems. The Windows image is created via scripted application installs and MDT task sequence. Both the Windows and Linux images are created via scripted installs, with MDT Task sequence for Windows and Kickstart for Linux. The images are then cloned by Nutanix image service into tiCrypt. Researchers can request applications that will be added to a custom image. Internet/network based licensing can be allowed through the tiCrypt proxy that denies all other internet access. The research VMs are further secured by FireEye Endpoint Security.

Table 4: Case Study - A multi institution collaborative longitudinal study

The Fragile Families and Child Wellbeing Study (FFCWS), a longitudinal study led by Princeton and Columbia Universities, follows participants from birth to 22 years of age with seven waves of data collection, including parent and child questionnaires, Geographic markers, DNA, Cardiovascular, Sleep, Brain, Administrative Data, and more. While the data is housed at Princeton, researchers from various institutions need access to the data to conduct research. The requirements for access to pre-release data among the Study's collaborative team of genomic researchers specify that approved researchers from other institutions should be able to get onto Princeton's systems to perform data analysis, but they should not be able to copy or move any of that data to their own systems. tiCrypt replaces a fragile configuration of Citrix that had met those requirements.[5].

3.3 Princeton's adoption of tiCrypt/Citadel

Researchers in the biomedical, social, and behavioral sciences are expected to design and perform research that ensures that research subjects' dignity, welfare, and privacy are protected, and appropriate information about individuals remains confidential. Additionally, federal, state, sponsor, and collaborator requirements (including DUAs with these entities) often drive the need for specific technological controls supporting research projects that ensure data is appropriately managed. Before establishing secure research infrastructure (SRI) called Citadel, Princeton University provided researchers with tailored, in-house solutions to address these data protection needs.

When surveyed on their experience with tailored solutions, researchers identified several pain points that led Princeton to strategize a new model. Starting a project required a long lead time, between 30-60 days, and the process for requesting SRI was unclear to the researchers. Once an SRI project was set up, the tailored solutions were reported to be cumbersome and hard to work with and as demands for the project changed, the system had to be manually resized. As each project required custom deployment and review by the Information Security Office, the model was not fit for replication and not scalable. To continue supporting the growing demand for SRI and the need to handle CUI, FISMA and HIPAA data, the Office of Information Technology sponsored a project to review available solutions that would meet this need.

The new SRI model needed to be easy to provision, ready to use with Linux and Windows operating system support and needed to allow pre-loaded research tools in the environment. Each project needed to be isolated with secure data egress and easy user access administration. Researchers wanted the projects to allow collaboration across institutions and to be simple to use. The SRI also needed to be NIST 800-171 (CUI), NIST 800-53 (FISMA) and HIPAA compliant, with comprehensive audit trails and to allow integration with OIT monitoring systems. Finally, the new model needed to be scalable to support growing demand for a multitude of projects ranging from small to large and complex.

Table 5: Princeton Vendor Assessment

Product	RightScale	ADRF	RHEDCloud	tiCrypt
Year Established	2010	2016	2019	2015
Vendor	Flexera	Coleridge Initiative, NYU	The RHEDCloud Foundation, Emory	Tera Insights, UF
Platform	Cloud Agnostic Management	AWS Gov-Cloud	Cloud Agnostic Controls	Private Cloud
Certifications	No	FedRamp moderate certified	No	CUI, FISMA
Availability	Within few months	Immediate	Depends on the cloud provider	Within few months
Customization	Yes, some	No, offered as a service	Yes	Yes
OS Availability	Windows Linux	Linux only	Windows Linux	Windows Linux
Supports Addition of Applications	Yes	No	Yes	Yes
Website	[10]	[3]	[9]	[12]

After reviewing Princeton's requirements, three critical factors weighed heavily in the selection of the solution:

- (1) Researchers must be able to securely analyze and store data in a user-friendly manner while preventing unauthorized access, change, and distribution enabling alignment with NIST 800-171 (CUI) regulations and standard DUA requirements.
- (2) The solution must position Princeton University to pass a design assessment of NIST 800-171 controls and, in the future, to achieve Department of Defense Cybersecurity Maturity Model Certification (CMMC) Level 3 certification.
- (3) Virtual computing environments must be available on-demand and scalable while minimizing the effort required to maintain the environment.

After a detailed review, tiCrypt was selected as it satisfied most of the user requirements, was reasonably priced, and was in a position to be operational in a short period of time. See Table 5 for details of the selection criteria.

For the Princeton implementation of tiCrypt, branded as Citadel, it was a priority to empower researchers with control over their projects and as such requested that an additional role, the sub-admin, be introduced to the system. Princeton worked closely with Tera Insights to define this role. By introducing a sub-admin into the hierarchy, Princeton is able to delegate user management, VM resizing, and data management to the researcher. Each project in Citadel has two sub-admins who can create VMs, control user management, scale the environment and determine where and what data to backup. By establishing sub-admin groups, researchers gain control and reduced response time for their project.

The team is also working closely with the ServiceNow group to create a form for requesting system access and to establish support protocols. The proposed form contains a table that defines projects with associated sub-admins. When the request is created by the researcher, the form will automatically route to the project's designated sub-admin. Training records and signed Rules of Behavior for the researcher are also included in the ServiceNow request. This will automate the request process and create an audit trail.

3.3.1 Princeton's Technical Implementation. At Princeton, tiCrypt is managed by the Research Computing Department within the Office of Information Technology (OIT)[8] and the Princeton Institute for Computational Science and Engineering (PICSciE)[7]. Research Computing chose to integrate tiCrypt into its existing framework of hardware and software environments to simplify integration efforts, maintenance and operation. The existing framework of operations can be broken down into the following principles:

- **Reliability:** Machines are purchased from vendors that provide enterprise grade hardware, tools and warranty for mission critical systems.
- **Migratability:** Vendor neutral technology is used whenever possible enabling flexibility in moving solutions from one system to another. For this reason Princeton uses Linux KVM for the VMs in tiCrypt over VMWare or other proprietary VM technologies.
- **Scalability:** Every system implemented must be scalable without requiring re-design of the components to support an increased demand for storage and computing in the SRI as is anticipated in the near future.

Table 6: Case Study - Video Ethnography Study

A team at Princeton's Office of Population Research seeks to address the question of "How do families support their children's early learning?" using highly innovative video ethnography through the New Jersey Families project. The team uses direct video observations of parent-child interactions to deepen the understanding of contextualized strategies that families from a wide variety of backgrounds and in different social and economic circumstances use to build skills in their young children. Over 463,000 video clips have been collected from in-home video recordings of 21 families in New Jersey. The data resides in Citadel and allows collaborators from various institutions to securely view and analyze the video data[6].

Princeton's tiCrypt implementation consists of three physical servers and two VMs:

- tiCrypt Backend and Storage: Dell EMC Enterprise PowerEdge R740xd2
- tiCrypt Virtual Machine Server: Dell EMC Enterprise PowerEdge R7525
- tiCrypt Audit and Systems Management node: Dell EMC Enterprise PowerEdge R640
- tiCrypt MultiFactorAuthentication node: VMWare Virtual Machine
- tiCrypt Mailbox node: VMWare Virtual Machine

These servers use Springdale Linux[11], a RedHat Enterprise Linux clone, as the operating system (maintained in house). Configuration Management is done through Puppet. The physical servers are located in Princeton's state-of-the-art High Performance Computing Research Center. To provide additional access control, the systems are separated from the rest of the infrastructure by a cage. Access is granted through a proximity card lock with automatic video recording.

As it relates to performance and security monitoring, Princeton continuously monitors all key performance indicators (KPI) with Zabbix. Data related to CPU load, Disc IO, memory and storage are stored and displayed on the Zabbix Server. Alerts and notifications are sent via Slack and email to systems administrators if any KPI exceeds the set limit. Integrating tiCrypt into Princeton's security information and event management (SIEM) allows the Security Operations Team to send alerts and reports if attempts are made to access the Systems from off-site, without authorization, or if modifications of the OS are made outside of the maintenance schedule. Moreover, the system is bound to Princeton's enterprise vulnerability scanner and monitored by Endpoint Protection Systems. This data is displayed in the Enterprise Status Dashboard of ServiceNow. Some performance indicators are forwarded via Nagios.

4 LESSONS LEARNED AND FUTURE PLANS

The University of Florida learned that custom designed systems for each research project is very labor intensive to operate across the expanse of smaller projects that required DFARS compliance. From the process of external review of the system by a third-party assessment organization (3PAO) UF learned that it is very important to have evidence for all the controls described as implemented in the SSP. Policies must exist in a location accessible to the users and researchers, procedures must be described in documents that can be read by those responsible for executing the procedures, and most importantly they must be the maintained master copies. Stale copies of policies and procedures are of no value.

To auditors the complexity and constraints of the academic world and its institutions are often unfamiliar. At times, they expect that controls and requirements can be implemented as they are in the private sector, where control over the entire organization is generally much stronger than in a university setting. To prepare for this scenario, it helps to select auditing companies that have experience working with universities. This increases the likelihood of a smooth audit that completes on time and on budget.

The research world is constantly changing as is its very nature. In the current state of cyber war, the compliance landscape is changing

more rapidly than ever, so when looking for an architecture for a secure system, it is important to look for the following criteria:

- (1) Does the system scale to large numbers of users? More work at academic institutions will be subject to data use agreements and safeguarding requirements. Whatever is built must be able to grow, and will most likely require a constant budget and staff. The main thing likely to be funded for growth is capacity.
- (2) Does the system support easy growth within any project? Projects often start small, but if they are successful, they are likely to involve growing data sets. Make sure that additional capacity can be provisioned without having to rebuild.
- (3) Does the system support complex collaborations? Even if a project starts out with only members from the primary institution as participants, later stages of these projects (as well as new projects) often see the need for external collaborators. Design the system and all compliance controls so that this need can be accommodated without a major re-architecture effort. [14][17]

At the University of Maryland, strong internal partnerships and support from senior leadership have been crucial to the implementation. Several large colleges that interact with restricted research data, primarily CUI, had a major stake in the environment and helped pilot the tiCrypt solution. The Office of the Provost and Division of Research assisted with funding for the positions and technology needed to create and operate the system.

Educating campus more widely about restricted research compliance obligations is ongoing. There is a close partnership with the Division of Research and with the Division of IT Security team on training and outreach events. It is highly recommended that partnerships across campus units are made to inform and instruct researchers about securing research data.

As use of the tiCrypt system has grown, new challenges continue to arise, specifically around researchers utilizing laboratory equipment and software that cannot be duplicated inside of the UMD enclave. The equipment is often running on legacy software due to the costs of upgrades, and thus does not have network access. Institutions with decentralized management of labs and technology should establish points of contact with campus labs and create plans around securing the resources. In some cases, contractual statements clarifying what data are CUI have helped avoid compliance issues. However as checks on compliance become more ubiquitous[15], requirements will necessitate bringing labs into compliance by leveraging tiCrypt.

Princeton University quickly learned that tiCrypt's primary model of strict user-based control over available resources would not satisfy the academic culture of their research community. To that extent, Princeton approached Tera Insights to add the sub-admin feature to tiCrypt. Sub-Admins can create VMs and drives, assign user to these resources and set permissions. This allows for distributed administration providing more autonomy to the researchers.

A new method of backing up the results of researcher work on the VM's was also critically important. Working with Tera Insights, a new feature was added to allow sub-admins to backup the drives attached to the VMs to outside storage provided by

the departments. The backup feature is strictly controlled by the sub-admin/researcher. These new features empower researchers to control the structure of their projects with the ability to independently add users and decide where and what to backup which reduces administrative hurdles allowing them to focus primarily on their work.

Finally, Princeton University learned that the right governance structure is crucial to the successful implementation of tiCrypt. Having all parties included in the decision making process ensured buy-in from the stakeholders from the start of the project. Without the researchers, the compliance offices, and the technology staff represented at the same table, it would have been difficult to successfully implement a secure research infrastructure that not only met all the project requirements, but also continues to facilitate research leveraging restricted data.

REFERENCES

- [1] 2018. *Health Insurance Portability and Accountability Act of 1996*. <https://www.cdc.gov/php/publications/topic/hipaa.html>
- [2] 2020. *Federal Student Aid: Protecting Student Information ? Compliance with CUI and GLBA*. Retrieved March 08, 2021 from <https://ifap.ed.gov/electronic-announcements/121820CybersecurityProtectStudentInfoComplianceCUIinGLBA>
- [3] 2021. *Coleridge Initiative, NYU*. <https://coleridgeinitiative.org/>
- [4] 2021. *Federal Information Security Modernization Act*. <https://www.cisa.gov/federal-information-security-modernization-act>
- [5] 2021. *Fragile Families*. <https://fragilefamilies.princeton.edu/>
- [6] 2021. *NJ Families Study*. <https://njfamilies.princeton.edu/index.html>
- [7] 2021. *Princeton Institute for Computational Science and Engineering*. <https://researchcomputing.princeton.edu/about/about-picscie>
- [8] 2021. *Princeton University Research Computing*. <https://researchcomputing.princeton.edu>
- [9] 2021. *RHEDCloud*. <https://www.rhedcloud.org/>
- [10] 2021. *RightScale*. <https://www.flexera.com>
- [11] 2021. *Springdale Linux*. <http://springdale.math.ias.edu/>
- [12] 2021. *Tera Insights*. <https://terainsights.com/>
- [13] 2021. *Tera Insights*. <https://ticrypt.com/>
- [14] 2021. *University of Florida Research Computing*. <https://www.rc.ufl.edu/>
- [15] Billy Mitchell. 2020. *GSA considering more CMMC-like cyber requirements in its governmentwide contracts*. Retrieved March 1 2021 from <https://www.fedscoop.com/gsa-cmmc-governmentwide-contracts/>
- [16] UMD Office of Research Administration. 2021. *Historical Data Office of Research Administration*. Retrieved March 1 2020 from <https://ora.umd.edu/data-reports/historic>
- [17] Gregory L. Orgill, Gordon W. Romney, Michael G. Bailey, and Paul M. Orgill. 2004. The Urgency for Effective User Privacy-Education to Counter Social Engineering Attacks on Secure Computer Systems. In *Proceedings of the 5th Conference on Information Technology Education* (Salt Lake City, UT, USA) (*CITC5 '04*). Association for Computing Machinery, New York, NY, USA, 177–181. <https://doi.org/10.1145/1029533.1029577>