中国科学: 信息科学 2020 年 第 50 卷 第 8 期: 1273–1274
SCIENTIA SINICA Informationis
《中国科学》杂志社
SCIENCE CHINA PRESS

人物风采

冯登国院士

1 经历简介

冯登国, 1965 年出生, 1995 年于西安电子科技大学获通信与信息系统专业博士学位, 博士学位论文获首届全国优秀博士学位论文, 同年进入中国科学技术大学研究生院 (北京) 博士后流动站工作, 1997 年入选中国科学院 “百人计划”, 2019 年当选中国科学院院士. 2000 至 2012 年担任信息安全国家重点实验室主任、国家计算机网络入侵防范中心主任. 现为中国科学院软件研究所研究员、博士生导师, 及密码科学技术国家重点实验室主任. 曾担任国家 973 计划项目首席科学家、国家 863 计划信息安全技术主题专家组组长、国家 863 计划信息技术领域专家组成员、国家信息化专家咨询委员会委员、全国信息安全标准化技术委员会副主任委员、中国密码学会副理事长等. 曾获得国家杰出青年科学基金、国家重点实验室计划先进个人、中国科学院青年科学家奖、中国科学院十大杰出青年等荣誉和称号.

中国科学院院士
冯登国

长期从事网络与信息安全研究工作, 在 *Theoretical Computer Science, Journal of Cryptology, IEEE Transactions on Information Theory, Crypto*等国内外重要期刊和会议上发表论文 200 多篇, 主持研究和制定国际和国家标准 20 多项, 获得国家发明专利 60 多项, 获得国家科技进步一等奖、国家技术发明二等奖等国家和省部级奖励 20 多次.

2 主要成就

主要学术成就涉及保密通信、网络安全和可信计算等方面.

一是面向国家安全重大战略需求, 创建了复杂环境下关键信息资源安全可信动态调度与可靠重构理论, 提出高可用动态安全通信体系结构, 解决了开放异构多域网络环境下保密通信的动态安全防护这一重大核心难题, 奠定了构建国家级保密通信系统的理论基础. 基于这些理论主持研制出国家重要保密通信系统, 首次形成了动态的国家级保密通信服务保障能力, 已成为国家重要部门的关键基础设施, 在国家重大活动、处突维稳、抢险救灾等工作中发挥了重大作用.

二是提出剩余类环上相关免疫函数的频谱特征刻画、多项式型 Bent 函数构造、可证安全的输入长度可变加密模式设计、条件掩码分析等新理论和新方法. 基于这些理论主持研制出密码算法综合检

测分析平台, 发明了我国首个成为国际主流标准的密码算法 – 祖冲之 (ZUC) 算法, 并得到广泛应用, 提升了我国在密码设计领域的国际学术地位, 对推动我国密码产业发展具有重大而深远的意义.

三是提出双层式入侵容忍证书认证系统结构模型和构造机理, 设计并实现了双层式入侵容忍证书认证协议, 有效克服了基于已有理论构建的 PKI (public key infrastructure) 在抗攻击性、可扩充性和可管理性等方面存在的缺陷, 为解决高安全等级 PKI 构建问题提供了全新的技术途径. 基于该理论主持研制出具有入侵容忍功能的高安全等级 PKI, 构建了我国 PKI 标准体系, 已应用于电子政务、电子商务等国家重要领域.

代表性论文著作

1 Feng D G. Three characterizations of correlation-immune functions over rings ZN. Theor Comput Sci, 1999, 226: 37–43
2 Hu H G, Feng D G. On quadratic bent functions in polynomial forms. IEEE Trans Inf Theor, 2007, 53: 2610–2615
3 Wang P, Feng D G, Wu W L. HCTR: a variable-input-length enciphering mode. In: CISC 2005. LNCS, 2005, 3822: 175–188
4 Zhang B, Feng D G. New guess-and-determine attack on the self-shrinking generator. In: ASIACRYPT 2006. LNCS, 2006, 4284: 54–68
5 Zhang B, Xu C, Feng D G. Practical cryptanalysis of bluetooth encryption with condition masking. J Cryptology, 2018, 31: 394–433
6 Feng D G, Pei D Y, Xiao G Z. Maximum correlation analysis of nonlinear combining functions. Sci China Ser E: Technol Sci, 1998, 41: 31–36
7 Feng D G, Xu J. A new client-to-client password-authenticated key agreement protocol. In: IWCC 2009. LNCS, 2009, 5557: 63–76
8 Chen X F, Feng D G. Direct anonymous attestation for next generation TPM. J Comput, 2008, 3: 43–50
9 冯登国. 频谱理论及其在密码学中的应用. 北京: 科学出版社, 2000
10 Feng D G, Xiang J. Experiences on intrusion tolerance distributed systems. In: 29th Annual International Computer Software and Applications Conference, Edinburgh, 2005

中国科学: 信息科学 2020 年 第 50 卷 第 8 期: 1127–1147
SCIENTIA SINICA Informationis
《中国科学》杂志社
SCIENCE CHINA PRESS
评述

创新发展中的可信计算理论与技术

冯登国[1,2], 刘敬彬[2,3], 秦宇[2*], 冯伟[2]

1. 中国科学院软件研究所计算机科学国家重点实验室, 北京 100190
2. 中国科学院软件研究所可信计算与信息保障实验室, 北京 100190
3. 中国科学院大学, 北京 100049
* 通信作者. E-mail: qin_yu@tca.iscas.ac.cn

收稿日期: 2020–04–20; 修回日期: 2020–05–29; 接受日期: 2020–06–02; 网络出版日期: 2020–08–03

国家重点研发计划 (批准号: 2018YFB0904900, 2018YFB0904903, 2020YFE0200600) 和国家自然科学基金 (批准号: 61872343, 61802375) 资助项目

摘要 可信计算以硬件安全机制为基础, 建立可信赖计算环境, 从体系结构上全面增强系统和网络信任, 是当前学术界和产业界的关注热点. 随着信息技术的深入发展, 新应用场景的不断涌现, 网络空间的安全威胁日益严峻, 因此可信计算在重要信息系统的安全防护领域将发挥越来越重要的作用. 本文从创新发展角度, 围绕作者 20 年来在可信计算领域的研究成果, 综述了可信计算理论的发展历程, 提炼总结出涵盖两大方法基础、三大信任核心和四大关键技术的可信计算技术体系, 阐述了移动可信计算、抗量子可信计算、可信物联网、可信云、可信区块链等方面的重要研究问题以及可信计算在这些领域的融合创新成果. 在移动可信计算方面, 软硬件结合的可信执行环境体系架构设计和实现是研究重点, 其次, 移动操作系统内核运行时安全隔离防护, 以及基于可信执行环境 (trusted execution environment, TEE) 的移动应用安全防护也是两个重要研究问题. 在可信物联网方面, 由于嵌入式环境本身的特性以及资源的受限, 轻量级的信任根构建、高效安全的软件证明、实用的安全代码更新机制、集群设备证明是该领域有待进一步研究的重要问题. 在抗量子可信计算、可信云、可信区块链等新型场景中, 可信计算技术也在不断地拓展其应用边界, 发挥更加重要的作用. 最后本文展望和讨论了可信计算未来的发展趋势.

关键词 可信计算, 可信执行环境, 移动可信计算, 抗量子可信计算, 可信物联网, 可信云, 可信区块链

1 可信计算背景和发展

随着移动互联网、物联网等新型计算环境的普及, 安全问题愈发严重, 尤其是恶意代码攻击严重威胁着用户的隐私和财产安全. 例如, 加密勒索病毒 WannaCry, 导致了大量用户的计算机无法正常使用. 熔断 (meltdown) 和幽灵 (spectre) 漏洞, 利用英特尔处理器中的设计缺陷, 严重影响计算平台的安

引用格式: 冯登国, 刘敬彬, 秦宇, 等. 创新发展中的可信计算理论与技术. 中国科学: 信息科学, 2020, 50: 1127–1147, doi: 10.1360/SSI-2020-0096
Feng D G, Liu J B, Qin Y, et al. Trusted computing theory and technology in innovation-driven development (in Chinese). Sci Sin Inform, 2020, 50: 1127–1147, doi: 10.1360/SSI-2020-0096

全性. 这些安全威胁主要利用计算平台上的安全漏洞进行攻击, 根本原因在于计算平台缺乏体系架构上的主动防御手段. 因此, 如何在体系架构上实现主动防御机制, 从底层芯片出发, 提供基于硬件的平台完整性和机密性保护的整体安全解决方案, 已经成为目前面临的根本问题.

可信计算的含义. 可信计算是一种主动防御技术. 它利用硬件属性作为信任根, 系统启动时逐层度量, 建立一种隔离执行的运行环境, 保障计算平台敏感操作的安全性, 从而实现对可信代码的保护. 可信计算可以实现对于攻击的主动免疫, 基于芯片中的硬件安全机制, 可以主动检测和抵御可能的攻击. 相对于传统的杀毒软件、防火墙等被动防御方式, 可信计算不仅可以在攻击发生后进行报警和查杀, 还可以在攻击发生之前就进行主动防御, 能够更系统更全面地抵御恶意攻击. 总之, 可信计算就是针对目前计算平台不能从根本上主动解决安全问题而提出的, 通过在计算平台中集成专用硬件模块建立信任锚点, 利用密码学机制建立信任链, 构建可信赖的计算环境, 使得从根本上解决计算平台的安全问题成为可能.

对于可信的定义, 不同的机构有不同的理解. ISO/IEC 将可信定义为[1]: 参与计算的组件、操作或过程在任意的条件下是可预测的, 并能够抵御病毒和一定程度的物理干扰. IEEE 将可信定义为[2]: 计算机系统所提供的服务的可信赖性是可论证的. 国际可信计算组织 (Trusted Computing Group, TCG) 将可信定义为[3]: 一个实体是可信的, 如果它的行为总是以预期的方式, 朝着预期的目标进行. TCG 的可信计算技术思路是通过在硬件平台上引入可信平台模块 (trusted platform module, TPM) 提高计算机系统的安全性, 这种技术思路目前得到了产业界的普遍认同. 我们的思路与 TCG 类似, 将可信理解为以安全芯片为基础, 依托安全硬件建立不受恶意代码攻击的可信执行环境, 确保系统实体按照预期的行为进行.

国内外研究现状. 早在 20 世纪 90 年代中期, 国外一些计算机厂商开始提出可信计算技术方案, 基于硬件密码模块和密码技术建立可信根、安全存储和信任链机制. 该技术思路于 1999 年逐步被 IT 产业界接受和认可, 形成可信计算平台联盟, 2003 年改组为 TCG , 并逐步建立起 TCG TPM 1.2 技术规范体系, 将其思路应用到计算机的各个领域, 并在 2009 年将该规范体系的 4 个核心标准推广为 ISO 国际标准.

2014 年 TCG 提出了 TPM 2.0 规范, 相比 TPM 1.2, TPM 2.0 支持更多的密码算法, 同时增加了授权层次, 具有更灵活的架构和更广泛的应用. 在产业界, 许多芯片公司都将部分可信计算功能集成到商用的处理器中, 如 ARM 公司的 TrustZone 技术、Intel 公司的 SGX 技术和 AMD 公司的 SEV (secure encrypted virtualization) 技术等, 都在处理器中实现了内存隔离, 可以为上层应用提供安全的执行环境, 保障敏感程序的安全性, 并被广泛应用在移动手机和云平台中.

我国也一直高度重视可信计算这一领域, 秉承着核心技术自主创新、信息安全自主掌控的理念, 积极推进可信计算的研究与发展, 颁布实施了《可信计算密码支撑平台功能与接口规范》[4]. 我国已经形成了基于可信密码模块 (trusted cryptographic module, TCM) 的可信计算技术体系, 并在移动可信计算、可信云等应用领域中实现了可信 Android 系统、可信工控系统、主动免疫嵌入式系统等, 在远程证明、可信执行环境核心技术研究上也取得了重要创新. 目前, 我国自主可信计算产业发展迅速, 在国家关键信息基础设施、重要信息系统中发挥着重要作用. 在安全启动方面, 安全 PC、服务器、平板电脑、智能手机、网络接入设备等已普遍集成 TCM, 完全支持基于 TCM 的安全启动和信任链构建. 在可信执行环境方面, 国产智能手机中已经普遍部署了支持指纹认证、人脸识别、电子支付等功能的可信执行环境安全应用, 这些产品正在通过不断的技术迭代提升其安全性. 在远程证明、可信存储等方面, 我国的技术实力也基本与国际 IT 企业处于同一水平. 总之, 无论是技术还是产业层面, 我国目前已经处在国际可信计算领域的前列.

可信计算面临的新挑战. 随着移动互联网、量子计算、物联网、云计算、区块链等技术的发展和应用, 可信计算技术也开辟了新的应用场景. (1) 在移动可信计算方面, 设计具有更小可信计算基 (trusted computing base, TCB) 的移动可信体系架构, 以及实现内核运行时和移动应用的安全防护是重要的研究问题. (2) 随着量子计算的发展, 设计高效的抗量子密码算法和协议是一个亟需解决的科学问题; 更进一步, 需要设计具有抗量子能力的 TPM/TCM, 并且构建抗量子可信计算技术体系. (3) 在可信物联网方面, 轻量级的信任根、高效安全的软件证明、安全代码更新机制是该领域的重要研究问题. (4) 在可信云中, 如何利用虚拟可信平台模块、虚拟机监控技术、新型的硬件安全技术实现云平台安全防护是该领域需要解决的重要安全问题. (5) 在可信区块链方面, 新型的可信执行环境技术可以为区块链提供新的思路, 例如, 利用硬件安全机制改进共识协议, 使用可信执行环境 (trusted execution environment, TEE) 保障区块链的计算环境等.

2 创新理论研究进展

虽然可信计算理论发展相对于产业应用而言并不瞩目, 但是它对于推动可信计算技术发展至关重要. 可信计算理论方法是可信计算技术、产品、标准的安全基础. 信任链、可信执行环境、可信云等技术的应用都依赖于密码算法和协议理论, 依赖于可信密码模块和可信软件. 近年来, 持续发展的可信计算理论一方面巩固着网络空间信任的基础, 另一方面也拓展延伸着新技术应用的信任边界.

2.1 可信计算理论发展历程

可信计算概念早在 20 世纪 80 年代就已出现, 主要用于可信计算机系统的安全评估, 重点关注如何保证计算机启动时和运行时的安全可信, 随后朝两个不同的方向发展: 一个是利用物理防篡改设备保障 TCB 可信, 以 TCB 为信任锚点构建计算机系统的可信计算体系; 另一个是构建隔离计算系统保障特定敏感软件代码运行环境的可信, 以此思路为演进, 发展出了基于 CPU 特殊安全模式的通用 TEE. 这两个发展方向的目标是一致的, 都是构建一个不受恶意代码攻击的可信赖的系统环境, 确保系统行为按照用户的预期目标进行, 只不过一个是通过固化信任成为不可篡改的物理芯片保障基础可信, 一个是通过隔离不可篡改的计算环境保障运行信任.

可信计算从固化信任根角度看发展脉络如下: (1) 最小 TCB. 自计算机诞生后, 最小 TCB 就是计算机系统最重要的安全设计目标之一, 最小 TCB 上实施最小权限分配可确保计算机最基本的安全控制. 可信计算机系统评估准则 (如 TCSEC, CC 等) 都将建立最小的 TCB 作为计算机安全的关键目标, 小型化 OS、可验证 OS、最小虚拟机监控器 (hypervisor) 等都尝试着构建小而安全、可用可控的 TCB. (2) 信任根. 信息技术发展出现了很多不同的硬件信任根, 如密码协处理器、防篡改硬件模块、增强型 CPU、智能卡、硬件令牌 (token) 等, 这些都植根于计算机硬件, 广泛应用于 PC、服务器和互联网环境. (3) TPM/TCM 通用信任根. 2001 年, 国际可信计算产业联盟 (TCG 前身) 推出了 TPM 技术标准, TPM 作为通用信任根广泛应用于 PC、笔记本电脑、服务器等. 我国于 2007 年 12 月发布了《可信计算密码支撑平台功能与接口规范》[4], 建立了通用的计算机信任根 TCM. TPM/TCM, 由于其通用性、绑定性、经济性, 在计算机产业界得到了广泛的应用, 全球主流的笔记本电脑、台式机上几乎都配置部署有这类安全芯片.

与固化信任根思路同时发展的是计算机系统的运行信任问题, 可信计算从运行信任的角度看发展脉络如下: (1) 专用可信系统. 这类系统着眼于解决计算机安全启动、计算机隔离执行等安全问题, 美国 20 世纪 90 年代的 Kent, Abyss, Citadel 系统就采用专用的防篡改硬件模块, 实现计算机的安全

启动和数据加密存储保护. 这类专用的可信系统的思路是隔离系统运行环境, 将应用系统分隔为两部分: 一部分运行于不受保护的通用主机环境; 另一部分运行于受保护的可信计算设备, 这部分可信子系统提供加密认证、文件安全存储、访问控制等基础安全服务. (2) 专用 TEE. 2000 年后的 XOM, MIT AEGIS 系统, 均采用与通用 X86 架构不兼容的增强改进型 CPU, 按照程序分隔成独立的运行隔间, 内存隔间中的代码和数据都是加密保护的, 只有可信的 CPU 才拥有解密密钥. 这些方案与专用可信系统相比具有更高的安全性, 安全假设更强, 其理论模型是假定应用程序不必信任操作系统和其他程序, 只需要信任 CPU. (3) 通用 TEE. 与专用 TEE 有完全相同安全模型和实施技术, 它针对通用的 PC 平台、智能手机平台等构建更为通用的 TEE, 近年来通用 TEE 成为了可信计算领域另一个重要发展方向. 通用 TEE 扩展通用 CPU 的安全功能, 在其特殊安全模式下增加内存隔离、数据代码加密及完整性保护等安全功能, 典型代表有 Intel SGX, ARM TrustZone 和 RISC-V Enclave. Intel SGX 的安全模型是不信任 Host 操作系统, TEE 只信任 Intel CPU, 主要应用于云计算服务器的安全应用. TrustZone 和 RISC-V Enclave 同样不信任 Host OS, 但除了信任 CPU 外, 还需要信任 Secure OS, 主要应用于移动嵌入式领域, 保护终端用户指纹、支付数据等. 这些通用 TEE 技术实现上都与 CPU 架构深度结合, 主要应用于云计算、移动互联网、物联网等新型应用的安全场景.

从可信计算的发展历程中, 我们可以看出, 可信平台模块 (TPM/TCM) 和 TEE 两个方面并行发展, 互为补充. 可信执行环境依赖于可信平台模块的支持, 而可信平台模块安全保护能力需要 TEE 扩展和增强. 固化信任与运行信任是相互关联的, 专用可信系统中的通用安全服务功能, 不断地归并集成, 最终小型化也会发展到通用的 TPM/TCM 信任根; 而最小 TCB 在通用 CPU 架构上通过不断扩展必要的基础安全功能, 就得到了通用 TEE. 可信计算的安全启动、度量证明、可信存储等安全特色在固化的信任根和通用 TEE 上得到了充分的应用.

2.2 可信计算理论自主创新发展

我国在可信计算理论研究方面几乎与国际研究机构、TCG 同时起步, 并且在算法、协议和安全性分析等方面取得了一系列重要且有影响的创新成果.

2.2.1 TCM 密码体系创新

早在 2004 年, 我国就开展了基于国产商用密码的可信计算密码方案的设计论证工作, 在可信计算密码研究方面取得了一些自主创新成果.

(1) 前瞻性、创新性地提出了对称密码与非对称密码相结合的密码方案. 2007 年我国发布了《可信计算密码支撑平台功能与接口规范》, 开启了自主设计可信计算技术和标准体系的先河, 自主可信计算体系与国际 TCG 技术标准体系架构相互兼容, 适用于中国自主密码技术应用体系, 但是 TCM 底层的算法和协议实现与 TCG 截然不同.

(2) TCM 除了对称密码应用方案和体系创新外, 在可信计算协议方面也进行了大胆的探索和创新, 提出了支持自主密码算法的 TCM 授权协议, 还有中国自主的直接匿名证明 (direct anonymous attestation, DAA)、远程证明等可信计算安全与隐私保护协议, 并支持中国特有的密钥协商、数字信封机制和双证书技术体系.

(3) TCM 密钥存储结构体系与 TCG 相比也具有较大的优势, TCM 内部存储密钥都是对称密钥, 与 TPM 1.2 的密钥存储结构有较大差异, 安全性和效率都有了较大提高. 在实际应用中, TCM 密钥加载性能与 TPM 1.2 相比至少提高 40%. 因此, TCM 应用具有较高的性能优势, 特别适用于高性能环境的服务器敏感数据保护.

由于我国可信计算学术界和产业界的不懈努力, 中国自主可信计算体系得到了国际社会和国际 IT 产业界的认可. 自 2012 年起, TCG 开始以密码技术本地化的方式公开支持中国密码算法. 在与 TCM 竞争发展中, TCG 也发现了 TPM 自身体系的不足, 借鉴 TCM 设计中的创新思路, 充分吸收了 TCM 密钥层次结构的创新, 提出了 TPM 2.0 技术标准.

2.2.2 直接匿名证明协议

DAA 是可信计算领域保护设备和用户隐私信息的重要协议, 这类协议可广泛应用于电子投票、在线拍卖、电子支付、医疗健康记录等匿名认证应用场景. 现有的 TPM/TCM 技术规范中都有对 DAA 的功能和接口的支持. 在可信计算远程证明中, 它用于 TPM/TCM 安全芯片及平台的匿名身份认证, 其优势在于匿名认证令牌 (秘密密钥) 存储在芯片内部, 受到严格的物理安全保护, 并且协议签名长度短, 计算效率高.

DAA 协议的研究经过了 RSA-DAA, ECC-DAA 两个协议更替的不同时期, 从 2004 年首个 DAA 协议提出, 到 TPM 2.0 规范、FIDO 技术规范采纳最新的 DAA 协议, DAA 协议经过长期的持续改进发展, 逐渐趋于完善.

2004 年国际上提出了首个 DAA 协议 BCC04 [5], 它建立了 DAA 协议基本安全模型, 并且给出了 DAA 协议可证明安全框架, 随后 TCG 的 TPM 1.2 芯片规范采纳了该 RSA-DAA 协议方案. 但是最初的 DAA 协议方案存在协议计算性能低的不足, 随后研究者们进行了一系列改进, 2008 年基于 LRSW 假设, 国际上提出首个基于 ECC 密码体制的 BCL08 方案 [6]. 紧接着我们提出了基于 q-SDH 假设的 ECDAA 协议 CF08 [7], 签名长度比 BCC04 缩短 90%, 比 BCL08 缩短 50%, 是当时签名长度最短的 DAA 协议, 进一步提升了 DAA 协议计算效率, 为后续 ECDAA 研究指引了方向, 也是目前 TPM 2.0 DAA 协议支持的代表性方案之一.

从 2010 年开始, ECDAA 协议安全模型趋于成熟, 协议设计和改进重点集中在减少芯片计算量. BL10 [8], CPS10 [9], BCL11 [10] 等协议对 DAA 协议的模型安全性、芯片计算量、预计算等进行了大量改进, 取得了很大进展. 2014 年, 我们在改进 BL10 的基础上, 提出了一个 TPM/TCM 计算量最小的基于 q-SDH 假设的 ECDAA 协议 SDH-DAA14 [11], 把安全芯片的计算量从 $3G_1$ 降低到了 $1G_1$. 同时还针对 TPM 2.0 技术规范草案中所采纳的 DAA 协议接口进行了安全分析 [12], 突破性地发现 TPM 2.0 规范在实现 DAA 协议的接口存在潜藏的 Static DH oracle 攻击, 攻击者可以多次调用 TPM 2.0 的 DAA 协议接口构造攻击, 将 DAA 协议的安全强度降低到 76 bits, 小于 80 bits 的基本安全强度. 在这些研究成果的基础上, 我们于 2014 年开始我国自主的 TCM ECDAA 协议标准的研究, 在原有 DAA 协议接口之上进行了优化改进, 支持更高性能的 DAA 协议实现, 最终完成了自主的可信计算直接匿名证明标准研制工作. 随着 TPM 2.0 ISO/IEC 11889 标准的发布, 尽管 ECDAA 协议已经完全成熟, 但是在后量子时代, 必然需要进一步推进抗量子的 DAA 协议的研究.

从可信计算 DAA 协议的研究历程中可以看出, 我们在关键研究节点作出了重要贡献. 我们提出和改进的 ECDAA 协议已经成为了 DAA 协议的代表性方案之一, 还突破性地发现 TPM 2.0 中 DAA 协议接口实现的缺陷, 增强了可信计算标准中 DAA 协议应用的安全性, 推动了可信计算理论的发展.

2.2.3 TPM 2.0 安全性评估

2012 年 TCG 发布了 TPM 2.0 技术标准草案, 在全球范围内推广下一代可信计算技术. 由于我国在国际可信计算标准升级中缺乏话语权, 必然会面对新的 TPM 2.0 技术是否足够可信安全、是否存在预设的标准漏洞等系列安全问题. 在这种背景下, 我们开展了 TPM 2.0 的安全性评估工作, 取得了一

系列重要的安全性分析和理论评估成果.

TPM 2.0 安全评估体系. TPM 2.0 评估首先需要建立一个完整的 TPM 2.0 安全评估体系. 在这个评估体系中, 评估对象为 TPM 2.0 芯片及其应用实现, 安全评估的基础是密码学方法、形式化方法, 以及自动化检测模型和工具 (采用自动化分析检测工具可显著提高 TPM 2.0 安全性分析效率) . 整项评估工作的目标有两个: 一个是 TCM 与 TPM 2.0 比较评估, 从架构兼容性、算法协议、安全功能、安全改进等角度评估二者的差异, 为我国 TCM 系列标准升级提供科学依据; 另一个是 TPM 2.0 安全性分析, 从 TPM 2.0 协议和接口分析、可信计算形式化分析的角度评估 TPM 2.0 技术标准的安全性, 并且找到标准中潜在的安全风险.

TCM 与 TPM 2.0 比较评估. TPM 2.0 规范草案支持我国密码算法作为可信计算技术实现方式之一, 原有的 TCM 安全功能作为 TPM 2.0 的功能子集, 这样 TCM 和 TPM 2.0 二者的差异性、兼容性评估就成为了重点关注的问题. 我们对 TPM 2.0 与我国自主 TCM 进行了全面的分析比较, 二者既有兼容也有差异: (1) 密码算法层面, TCM 与 TPM 2.0 二者完全兼容, 即 TPM 2.0 可支持实现 TCM 算法层面上的全部应用; (2) 安全功能层面, TCM 是 TPM 2.0 的一个功能子集; (3) 在数字信封、双证书体系、密钥存储结构等方面, TCM 与 TPM 2.0 存在较大差异, 二者并不兼容; (4) TPM 2.0 新增的安全机制存在一定的安全风险, 我们的理论研究 [13] 表明, TPM 2.0 需要进一步评估检测.

TPM 2.0 安全功能分析. 我们对 TPM 2.0 中近 20 项安全功能进行了全面的安全性分析和评测, 特别是 TPM 2.0 中新增的密钥协商、策略授权等安全功能. 密钥迁移方面, 我们对 TPM 2.0 的密钥迁移进行了安全性分析, 采用基于标记转换系统的形式化分析方法对密钥迁移建模, 利用 Tarmarin 自动化分析工具进行检测分析, 发现 TPM 2.0 密钥迁移存在导出密钥攻击 [14]. 密钥存储安全方面, 采用类型系统的方法对密钥存储保护进行分析, 利用定理证明方法证明了 TPM 2.0 密钥层次保护功能是安全的 [15]. 密钥协商方面, 我们在 CK 模型下分析了 TPM 2.0 的两阶段密钥协商 API 接口, 发现密钥协商存在一定的安全隐患, 攻击者可利用暴露的中间 Z 值实施 UKS (unknown key share) 和 KCI (key-compromise impersonation) 攻击 [16]. HMAC 授权会话方面, 我们在计算模型下使用安全协议验证工具 CryptoVerif 分析了 TPM 2.0 的 HMAC 授权, 证明 HMAC 授权会话满足认证性 [17]. 策略授权方面, 采用应用 Pi 演算形式化分析了 TPM 2.0 的策略授权安全机制, 通过 Tarmarin 自动化验证工具我们发现 NV 策略授权存在 TOCTOU 攻击 [18]. DAA 协议接口方面, 我们发现 TPM 2.0 规范所实现的 DAA 协议 API 存在 Static DH oracle 攻击的风险 [19], 并且基于进程代数对 DAA 协议实现建模分析, 利用进程演算验证工具 ProVerif 自动化分析发现 DAA 协议不满足前向匿名性.

这些 TPM 2.0 的安全性分析研究受到了国内外专家学者的关注, 我们把 TPM 2.0 规范的部分评估结果反馈给了 TCG, ISO/IEC 等国际标准组织. 例如, TPM2_PolicyNV 可能遭受 TOCTOU 攻击, TCG 接受了我们提出的评估意见并在 TPM 2.0 国际标准中进行了修订. 此外, 欧洲 Future TPM 项目 TPM 模型报告 [20] 中对我们的 TPM 2.0 安全分析成果给予了充分的肯定. 这些工作不仅提高了我国可信计算理论的研究水平, 而且促进了国际可信计算技术标准的提高和完善, 也是创新发展中的中国可信计算对国际可信计算的重要贡献.

3 可信计算技术体系

可信计算从最初的 PC 平台, 逐渐扩展应用到云计算、移动互联网、物联网等领域, 虽然可信计算的核心理念没有变, 但其技术形态已发生了巨大变化. 安全 PC 和服务器等传统可信计算平台侧重于主机的硬件信任根和信任链; 云计算则要求保障用户代码、数据的机密性和完整性, 可信计算平台侧

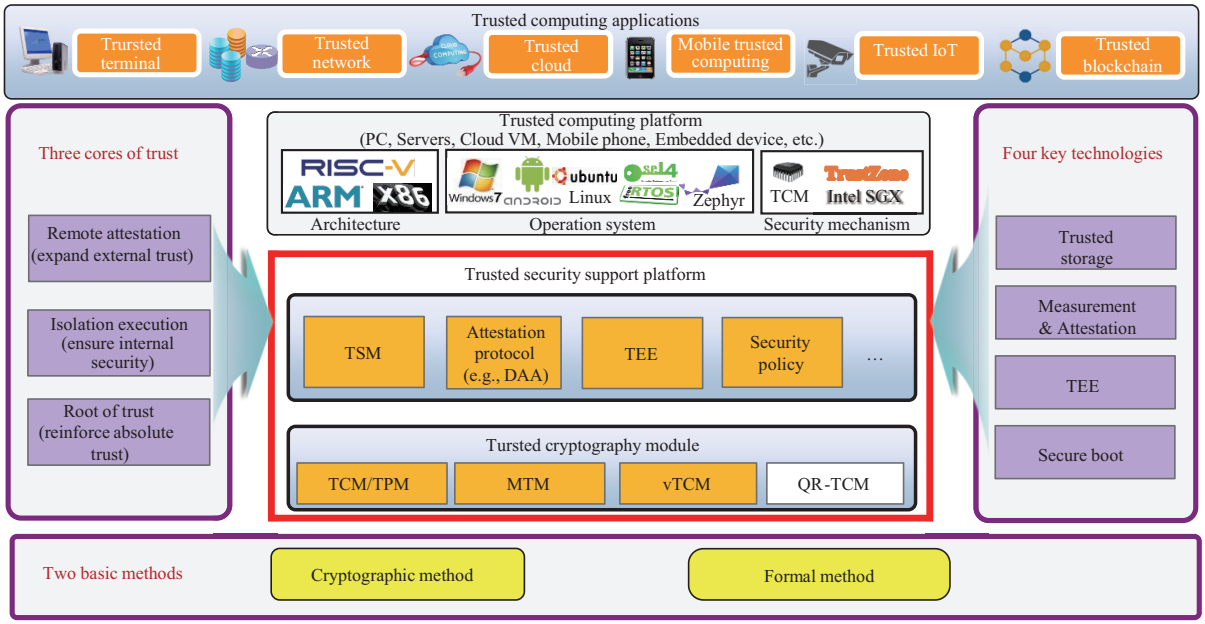

图 1 (网络版彩图) 可信计算技术体系

Figure 1 (Color online) Technology framework of trusted computing

重于可信执行环境和隐私保护技术; 物联网需要基础的设备认证和软件证明, 可信计算平台则倾向于轻量级的信任体系. 在这一背景下, 我们根据通用可信计算的技术特点, 提出了应用驱动的、可安全扩展的可信计算技术体系 (如图 1 所示). 其中密码学方法和形式化方法是理论基础, 可信密码模块是信任根, 可信安全支撑平台是核心. 可信安全支撑平台包含多种不同模式的信任根, 是可信计算平台最重要的系统安全组件, 它构建与主系统相互隔离的可信执行环境, 依据安全策略强制实施系统的安全防护, 并且对外提供密码计算、远程证明等可信计算服务, 具备安全免疫、主动防御的系统安全保障能力.

总体上看, 可信计算技术体系中包含两大方法基础、三大信任核心、四大关键技术.

两大方法基础. 两大方法基础是密码学方法和形式化方法. 可信计算中的所有算法、授权协议、远程证明协议、应用协议, 还有数据机密性、代码完整性、设备认证性等都是依靠密码学方法所保证的. 而形式化方法主要用来分析验证可信计算平台 TCB 的安全性, 特别是采用自动化分析工具, 分析评估隔离环境的 TEE 安全操作系统和可信应用代码的安全性, 减少潜在的软件风险. 总之, 密码学方法保障可信, 形式化方法分析验证安全. 密码学方法从算法和协议角度保障芯片、服务和应用的可信性, 而形式化方法从软件安全的角度确保隔离系统自身软件和应用的安全性.

三大信任核心. 可信计算技术体系的三大信任核心是信任根、隔离执行和远程证明, 它们从 3 个层次构建了可信计算平台的信任: 信任根固化绝对可信、隔离执行保障内部安全和远程证明拓展外部信任. 第 1 层信任是在物理层的信任根, 它是将权威机构、企业等用户信任锚点固化在系统中的最基础的信任, 是一种由权威可信第三方担保并且提供认证证据的绝对可信, 或者是无条件可信. 如果以嵌入在主板上的 TPM/TCM 作为硬件信任根, 那么信任根就是受物理防篡改保护的最小安全功能组件, 具有主动免疫的特性. 第 2 层信任是在系统层的隔离执行, 信任根不具备应用软件代码保护能力, 因而需要一个基于 CPU 的隔离受保护运行环境保障应用代码、机密数据的安全性. TEE 是实现隔离执行的重要技术手段, 它保障了系统隔离、代码验证和策略实施. TEE 是一个独立于主系统之外的安全

隔离子系统, 它上面运行的代码不会受到主系统的影响, 即使主系统被病毒、木马等攻陷, TEE 仍然是安全的. 结合内存页加密、地址随机化、数据执行保护 (data execution prevention, DEP)、影子栈等安全增强机制, TEE 还能够提供更强的系统安全性. 代码验证就是由于 TEE 内部的软件代码数量有限, 能够采用形式化方法证明其安全性, 减少代码潜在的漏洞和风险. 策略实施是 TEE 对主系统强制执行的安全控制机制, TEE 可以对主系统的敏感操作进行安全检查, 可以依据安全策略检测主系统内核和应用程序状态, 并且对外证明当前系统的可信性. 第 3 层信任是网络层的远程证明, 它是可信计算特有的一种安全机制, 将可信计算平台的内部信任通过网络验证拓展到外部. 传统的检测 (detection) 机制是在系统外部检查系统的运行状态是否正常, 而可信计算的远程证明机制是系统基于信任根或可信执行环境主动向外部证明自身运行状态, 是一种主动的、可验证的安全机制.

四大关键技术. 可信计算技术体系的四大关键技术是安全启动、可信执行环境、度量与证明、可信存储. 四大关键技术不仅要支撑可信计算技术体系的三大信任核心的构建, 而且要从各个角度保障可信计算平台的整体安全. 安全启动是构建可信计算平台信任的第一步, 确保系统的初始可信; 可信执行环境建立隔离受保护的安全计算环境, 确保系统运行时信任; 度量与证明是基于信任根或可信执行环境度量系统当前运行状态, 并且对外认证当前系统和软件可信, 确保信任的网络延伸; 可信存储是从数据安全的角度, 保障系统运行时的各类敏感数据的机密性和完整性, 确保系统的数据存储安全.

4 可信计算核心技术的融合创新

随着云计算、5G/6G、物联网、人工智能等新技术和新应用场景的出现, 可信计算正在多个领域快速融合发展, 为新技术应用提供强有力的安全支撑. 下面我们将从移动可信计算、抗量子可信计算、可信物联网、可信云、可信区块链 5 个方面对可信计算技术融合发展的最新进展进行详细的阐述.

4.1 移动可信计算

随着移动支付、指纹等功能在移动智能手机的普遍应用, 对这些移动应用和敏感数据的安全保护已成为移动互联网时代的共识, 于是, 移动可信计算应运而生, 它通过硬件特性为移动终端设备建立芯片级的安全防护, 从体系架构的层面保障用户设备的安全. 下面介绍移动可信计算国内外和我们的一些研究进展.

4.1.1 移动可信计算研究背景

与传统的恶意软件防护、APP 加固等通用技术不同, 移动可信计算是从移动终端设备系统信任的角度, 建立芯片级和 TEE 级的安全防护体系. 它通过构建 TEE, 实现平台完整性、安全存储、隔离执行、设备证明和远程配置等安全功能.

移动可信计算现已推出一系列标准: NIST 800.164、GlobalPlatform 的 TEE 标准、TCG 规范 (包括 TPM Mobile, MTM, TPM 2.0)、各种移动 HSM APIs (如 JSR 177, 类智能卡密码接口)、PKCS 11 等. 以这些标准为基础, 工业界已有一些 TEE 主要解决方案, 包括 ARM TrustZone 技术、Intel SGX 技术、安全元素方案 (secure elements, SE) 等. 这些方案中, SE 的安全性最高, SGX 次之, TrustZone 相对较低. SE 可以抵御大部分物理攻击和软件攻击, 主要应用在智能卡、部分手机中; SGX 可以抵御内存攻击和软件攻击, 无法抵御侧信道攻击, 安全性次之, 主要应用在高性能 PC 和云环境中; TrustZone 只能抵御软件攻击, 成本最经济但安全性相对低, 主要应用在移动和嵌入式领域.

TEE 目前已经在产业界获得了广泛应用, 其中基于 TrustZone 的方案是目前最流行的移动安全支

撑方案. TrustOTP [21] 提出了一个使用智能手机作为安全 OTP (one time password) 生成器的安全方案, 并实现了轻量级的可信 I/O 路径. 我们基于 SRAM PUF 的平台信任根和 TrustZone 技术给出了一个智能手机双因子方案 [22], 相比于专门的硬件安全令牌, 具有成本低、高效、灵活等优势. 微软学者 [23] 为 TEE 技术提出了一种可信 I/O 设备组件的软件抽象模型, 该模型定义了可证明的可信 I/O 路径, 并分别利用 Credohypervisor 和 TrustZone 技术在 PC 和移动平台上给出了原型实现. 其他方案还有安全手机支付方案 [24]、可信语言运行时架构 [25]、轻量级 Web 浏览器安全内核 VerifyUI [26]、安全广告内容验证模块 AdAttester [27].

4.1.2 移动可信计算体系结构

目前移动可信计算架构的研究已有许多成熟的方案. 根据其平台不同, 可以划分为基于 Intel 平台、ARM 平台和 RISC-V 平台. (1) Intel 平台上的架构方案研究较多, 其中最典型的是卡耐基梅隆大学 (Carnegie Mellon University) Cylab 实验室的 Flicker [28]. 该方案构建了独立于 OS 的轻量级 TCB, 构建从底层核心组件到上层应用的动态信任链, 为应用程序的敏感代码片段提供隔离保护, 使得平台状态证明不再依赖 OS 是否可信. 然而 Flicker 对系统的运行效率影响较大, 对于一些对效率要求较高的小型工作任务并不适用. (2) 基于 ARM 平台的主要方案有 Graz 大学 IAIK 研究机构提出的方案 [29], 其将一个轻量级 Linux 内核植入 TrustZone 安全世界, 构建了基础的 TEE 运行环境并实现了部分可信计算功能; 以及 Open Virtualization 推出的 SierraTEE [30], 其支持 OpenSSL 等密码学库以及第三方程序加载、文件系统等基本的系统服务. (3) 基于 RISC-V 平台的架构主要有麻省理工学院提出的 Sanctum [31], 在实现类似 Intel SGX 功能的基础上, 加入了对侧信道敌手的防御. 加州伯克利大学 (University of California, Berkeley) 提出的 Keystone [32], 通过在不可信的组件 (如操作系统) 下加入一种基于可编程中间层实现的内存隔离机制建立可定制化的 TEE. 虽然不同方案在 TCB、性能和功能等方面差异较大, 但这些方案都是基于现有的一些硬件安全特性来实现对软件敌手和侧信道敌手的防御, 可以保证 TEE 中可信应用与不可信系统的隔离, 提供较强的安全性, 但一般都缺乏对物理攻击的防御.

我们针对 TrustZone 系统内存数据缺少机密性和完整性保护的问题, 提出了一种基于软件的移动可信计算体系架构 SecTEE [33], 能够抵抗板级物理攻击. SecTEE 主要基于 CPU 的隔离机制, 并不要求特定的安全硬件, 如不需要 CPU 中具有内存加密引擎. SecTEE 与基于硬件的安全飞地 (secure enclave) 具有相同等级的安全性, 可以抵抗特权软件攻击、板级物理攻击, 以及侧信道攻击. SecTEE 构建了一个为 Enclave 程序提供安全隔离和功能保障的 TEE 系统, 它利用 SoC-bound 执行环境和隔离机制实现了物理和软件安全, 并且在内核提供了可以抵抗侧信道攻击的机制. 此外, SecTEE 还提供了丰富的可信计算功能.

4.1.3 可信安卓系统

安卓系统是目前移动平台上主要的操作系统之一, 广泛应用于移动智能手机设备中. 但是安卓系统的安全漏洞层出不穷, 缺乏从硬件层面提供的安全防护, 并且也不能提供可信计算的功能. 我们采用可信计算技术实现了可信安卓解决方案, 该方案以自主密码算法和 TCM 芯片级安全硬件为基础, 建立安卓镜像可信启动、系统运行环境可信监控、磁盘分区加密和验证等安全机制, 实现了对安卓系统启动、运行时以及长期存储设备的全方位保护, 并应用到智能手机、平板电脑等产品中. 可信安卓系统基于 TEE 的可信计算技术, 采用自主密码算法和 OS 度量构建 Android 静态信任链, 从代码和数据两个维度增强 Android 系统信任, 并结合 SEAndroid 技术和代码签名机制, 保障 APP 的代码完整性.

4.1.4 基于 TEE 的可信软件保护方法

基于 TEE 的可信应用程序安全防护. 可信应用程序安全防护, 也就是解决不可信操作系统运行可信应用程序的安全问题, 这是目前移动可信研究的一个重要方向. 随着 TEE 中功能的增多, 可信计算基也会变大, 这样 TEE 的攻击面也会相应增大. 目前移动可信应用程序防护主要的实现方式有两种: 一种是利用 TrustZone 本身的安全特性, 动态增加安全应用; 另一种是利用缓存作为安全的执行环境. 第 1 种方式的代表性方案是 TrustICE [34]. 该方案通过动态改变 TrustZone 安全世界的边界范围为普通世界的应用提供隔离保护, 解决了现有 TrustZone 方案 TCB 过大的问题, 其思路是 TCB 只负责核心安全功能, 同时保障安全性和开放性. PrivateZone [35] 创建了一个介于 TEE 和 REE 之间的执行环境 PrEE, 它利用新的执行环境运行敏感应用, 通过对页表进行重写映射确保了 PrEE 的安全性, 解决了传统 TrustZone 方案在安全性和开放性方面的冲突. 第 2 种方式的代表性方案是 CaSE [36], 该方案使用 TrustZone 和缓存作为 RAM 的技术创建了一个基于缓存的执行环境, 解决了 TrustZone 无法抵御冷启动攻击的问题. 其他方案还有 Ginseng [37], 其思路与前者类似. 现有的这些方案可以解决 TCB 过大的问题, 但并没有考虑受保护应用如何安全地调用不可信 REE 系统服务的问题, 实用性方面存在较大缺陷.

针对 TEE 应用粗粒度隔离而无法抵抗内部域攻击的问题, 我们在 TEE 安全架构上提出了一种基于安全域划分的应用程序内部强隔离保护方法 [38], 该方法具有更小的 TCB, 并且能够有效地抵抗 ROP, Iago 等内核攻击. 该方法改变了传统 TrustZone 隔离架构, 将需要保护的敏感应用移至普通世界, 大大减小系统的 TCB. 同时, 安全世界对普通世界系统底层行为进行拦截监控, 确保提供与安全世界等价的隔离保护.

运行时安全防护和内核隔离. 实现运行时安全防护和内核隔离是移动可信的另一个重要方向. 由于 TEE 具有较高的安全性, 其中一种研究思路是使用 TEE 作为系统监控的环境, 用来保障 REE 中系统和敏感应用的安全性. 它通过在 TEE 中实现一个系统监控的程序, 可以保障 REE 中程序运行时的安全性, 当程序遭到篡改和破坏时可以及时发现, 并进行报告和应急处理. 这个研究方向的主要思路根据保护对象的不同可以分为保护内核状态的方案与保护外设数据的方案. 保护内核状态的代表性方案是 TZ-RKP [39]. TZ-RKP 利用一个安全监控器来监控内核防止被篡改, 并创新性地将 REE 中控制特定系统函数的功能剥离, 而改为由 TEE 来进行监控和处理, 这种方式可以有效地阻止篡改内核的攻击, 实现完善的内核监控, 相对于周期性的监控做出了巨大的改进, 但其缺点是效率相对较低. SPROBES [40] 方案使用对特定指令进行中断的方式进行监控, 并且使用不变量的方式进行了保护. 其他代表性方案还有 TrustDump [41] 等. 保护外设数据的代表性方案有德国 TU Darmstadt 提出的方案 [42], 该方案通过识别外界环境采取不同安全策略限制 I/O 设备的行为, 解决了受限环境的个人移动设备使用的问题, 其优点是具备细粒度的策略控制能力, 可以做到对恶意代码和恶意行为的审查和阻止. 其他方案还有 ANDIX [43], Free-TEE [44] 等. 现有的 TEE 系统监控技术需要在内核层插入监控逻辑, 通过与 REE 系统组件的交互获取有效的系统状态信息. 而目前尚不存在高效完善的对 REE 监控代码本身进行保护的机制, 导致 REE 内核攻击者可能伪造状态信息, 篡改监控代码或直接绕过 TEE 的控制.

针对移动 OS 内核数据易被 root 攻击和篡改, 以及现有安全防护方案安全性不高的问题, 我们采用运行时保护和内核监控技术, 提出了 TEE 动态监控普通 OS 内核的安全体系架构 [45], 能够有效防止内核恶意代码注入和关键数据篡改. 该方案在不可信操作系统之上为合法进程提供可执行文件、运行时代码和控制流 3 个层次的完整性保护, 确保目标设备只能执行符合安全策略的授权代码.

4.2 抗量子可信计算

一旦通用量子计算机问世, 会对传统密码的安全产生重大影响, 波及到现有可信计算体系的安全性, 目前国际上在现有抗量子密码算法和协议基础上提出了基于抗量子 TPM 的一系列解决方案.

4.2.1 抗量子可信计算研究背景和现状

随着量子计算理论的发展, 部分经典模型下的计算困难问题可以在量子计算模型下有效求解, 现有基于传统困难问题的密码算法和协议将面临严峻的挑战. 2019 年谷歌声称, 在具有 53 个量子比特位的量子计算机上约 200 s 完成的任务在传统计算机上需要执行一万年, 显示出量子计算机相比于传统图灵计算机根本性的优势. 1994 年 Shor [46] 提出了可以在多项式时间内有效求解大整数分解问题和离散对数问题的量子计算模型, 使得工业界广泛使用的公钥密码算法与密钥交换协议从理论上变得不再安全. 对称密码算法和杂凑算法也受到了 Grover 等量子搜索算法的影响, 但相对容易解决, 主要通过增加密钥规模和杂凑值长度解决.

量子计算机的威胁大大促进了抗量子密码算法的研究与发展. 目前, 世界上已有诸多研究机构开始对抗量子 (quantum-resistant) 密码学进行研究. 2015 年 NIST 举行后量子安全研讨会, 并于 2017 年开始征集后量子密码算法, 目前已进入第 2 轮评估阶段.

目前被认为具有抗量子能力的公钥密码主要有: 基于杂凑的密码算法、基于编码的密码算法、基于格的密码算法和基于多元二次方程的密码算法等. NIST 认为, XMSS 和 LMS 等带状态的基于杂凑的签名算法能够抵抗一定强度的量子攻击, 并且此类算法除密钥生成需消耗较长时间外, 其签名和验签效率与现有的签名算法相当.

抗量子密码学研究的另一个重点是抗量子安全协议, 在可信计算协议研究中, 基于格的 (lattice-based) 抗量子直接匿名证明是一个研究热点. 当前已提出了基于格的直接匿名证明协议, 而后又有针对 TPM 环境的基于格的直接匿名证明协议 (L-DAA) [47] 被提出. 目前基于格的困难问题被认为具有抗量子攻击能力, 因此, L-DAA 有望成为下一代直接匿名证明协议.

目前国际上对抗量子可信计算平台模块的研究, 以 FutureTPM 项目最具代表性. FutureTPM 是欧盟发起的一项前沿性研究项目, 旨在设计和开发具有抗量子能力的可信计算平台模块. 其主要目标是实现从现有广泛应用的 TPM 系统到未来具有抗量子功能的系统的平稳过渡.

4.2.2 抗量子可信计算技术体系

抗量子可信计算技术体系在架构上与传统可信计算技术体系类似, 不同之处在于抗量子可信计算技术体系需要以抗量子密码算法和协议实现相应的可信计算平台功能. 此外, 还需要考虑到传统体系的兼容性以及更高需求的计算性能.

我们将抗量子密码算法和协议集成到可信密码模块中, 为上层设备提供抗量子计算的信任根, 形成抗量子计算的身份认证、远程证明、数据保护等可信计算机制; 同时研究了可信软件服务体系构建方法, 构建从内核层到应用层的可信软件服务体系, 为设备调用可信计算服务、管理资源提供了技术支撑. 在此基础上, 我们建立了抗量子可信计算技术体系架构, 主要内容包括: 建立了面向可信计算应用的抗量子安全模型; 提出了抗量子的可信计算协议可证明安全规约方法; 提出了基于抗量子密码算法的可信密码模块设计; 构建了支持抗量子可信密码模块应用通用接口的可信软件服务体系.

由于使用了抗量子密码算法、协议和接口, 将会导致抗量子可信密码模块和现有可信计算软硬件体系的兼容性问题. 例如, 主板 BIOS, Bootloader 的信任链因算法和协议的不同无法与抗量子可信密码模块集成. 采用抗量子密码算法后, 芯片上的密码学运算、度量证明的效率也是一大问题: 受限于

可信密码模块的运算速度、缓冲区大小等硬件条件, 大多数抗量子密码算法的运行速度将显著低于现有规模化商用的 RSA、椭圆曲线等密码算法. 受密码算法改变影响, 同时增长的还有度量日志的长度, 可能会导致新的缓存和存储问题. 此外, 基于 TPM 2.0 规范的一些安全机制 (如依赖加密算法的增强授权等功能) 还需要进一步考虑兼容性问题.

4.3 可信物联网

可信物联网是将可信计算技术应用于物联网的终端、网络以及云端来确保整个物联网的安全可信, 是物联网安全研究中的一个重要方向. 下面从轻量级物联网信任的角度对可信物联网的研究现状、研究问题和思路进行综述, 同时介绍我们在轻量级信任根体系、软件证明、安全代码更新等方面的部分研究成果.

4.3.1 轻量级信任根体系

在物联网场景中, 存在很多资源受限的嵌入式设备. 因此, 需要针对物联网设备本身的特征设计更加轻量化的解决方案, 提供适度有限的可信计算功能.

目前的轻量级信任体系构建方法可以归纳为以下 3 种: (1) 最小化硬件信任根. 该方法专注使用最小的硬件修改为物联网嵌入式设备提供基本的隔离执行和证明能力. 典型代表是 SMART 机制 [48]. (2) 基于 CPU 的应用保护. 该方法基于主流微处理器扩展其 CPU 指令以支持可信安全功能. 典型代表是 Sancus 机制 [49]. 最近, ARM 提出的平台安全体系 (ARM PSA), 将芯片安全扩展 TrustZone-M 技术也引入了小型嵌入式设备. (3) 细粒度内存访问控制. 该方法主要基于嵌入式设备本身具备的内存保护机制 (如 MPU) , 实现一个通用的嵌入式可信安全体系. SPM 方法 [50]、TrustLite [51] 和 TyTAN [52] 是其中的代表性方案.

我们针对功能单一的嵌入式设备提出了一种基于 PUF 的轻量级信任构建方案 [53]. 该方案采用物理元器件 SRAM PUF 的鲁棒性、不可克隆、不可预测等良好安全特性, 实现嵌入式平台的信任根密钥派生, 不需要额外的安全密钥存储, 具备标识唯一性和厂商可定制性. 此外, SRAM PUF 还可以作为熵源, 为密码算法和协议构造高安全随机数. 使用 SRAM PUF 派生的设备密钥作为根密钥, 并以 SRAM PUF 生成的高安全随机数为基础, 构建密钥存储体系, 我们实现了物联网计算环境的安全存储、加密封装、远程证明等可信安全功能. 与安全芯片、TrustZone、安全元素 SE 等重量级方法相比, SRAM PUF 信任根不需要改造硬件体系、可以灵活更新密钥、容易与轻量级密码算法融合、按需提供最小化可信计算功能, 因此, 更加适合资源受限的物联网设备.

4.3.2 软件证明机制

软件证明是一种在计算能力、系统资源等受限的环境下确保软件可信性的方法. 软件证明的目标是确保嵌入式设备上运行的软件可信, 防止软件篡改、恶意代码注入等.

由于软件证明缺少安全硬件, 其安全性相比基于硬件的远程证明弱一些, 容易遭受各种攻击, 如内存替换或内存复制攻击、内存压缩攻击、代理攻击、SPLIT-TLB 攻击 [54] 等. 针对这些安全问题, 已提出一些特殊的软件构造以增加攻击的难度, 提升软件证明本身的安全性. 软件证明方法可以归纳为以下 3 类: (1) 基于时间的证明. 主要思想是对设备的整个程序内存进行校验和计算, 如采用伪随机内存遍历、迭代运算等方式保证任何攻击行为都会明显影响计算过程, 典型代表有 SWATT [55], SBAP [56], VIPER [57] 等. (2) 基于内存随机填充. 使用不可压缩的伪随机噪声对嵌入式设备空闲的内存区域进行填充, 并将这些伪随机值也包含在校验和的计算过程中, 使得攻击者无法利用空闲的内存空间, 典

型代表有 Yang [58], Memory-printing [59] 等. (3) 基于证明函数随机构造. 前述方法的证明函数基本都是固定的, 容易被离线静态分析和逆向工程破解. 该类方法采取了一种新思路, 外部验证者在每次证明协议时, 先构造一个随机的证明函数, 并将该函数 (代替随机挑战值) 发送给证明者设备, 从而避免逆向攻击, 其典型代表为 PIV [60].

通用的软件证明主要是静态机制, 而且大部分针对单个证明者的场景. 这存在两方面的局限: 一方面是静态证明主要度量设备上的程序代码, 无法防止运行时控制流等攻击, 如 ROP 攻击; 另一方面物联网通常包含海量异构节点, 单个证明者的证明协议很难满足大规模设备认证的需求, 需要更加高效灵活的集群设备证明. 因此, 目前软件证明有两个热点研究方向: (1) 控制流证明. 这种证明不仅考虑程序代码的完整性, 同时还考虑控制流的完整性. 典型代表是 C-FLAT [61] 和 LO-FAT [62], 它们通过在二进制级别对程序的执行路径进行度量, 弥补了静态证明的不足. (2) 集群证明. 这种证明方法适合海量互联设备, 兼顾到证明方法的可扩展性、通信以及计算复杂度. 典型代表是 SEDA [63] 和 SANA [64].

在软件证明方面, 我们完成了 3 方面的工作: 一是针对通用物联网嵌入式设备, 基于 PUF 轻量级信任根设计了支持双向认证的高安全软件证明方案 AAoT (attestation and authentication of things) [65], 结合 PUF 硬件和软件证明的优势, 能同时满足嵌入式设备的身份认证和完整性证明安全要求, 属于软硬件结合的轻量级构建方法; 二是针对物联网集群环境, 提出了一种基于设备分组的高效集群证明方法, 解决了传统软件证明方法可扩展性问题; 三是为了应对 ROP (return-oriented programming) 等运行时攻击难题, 提出了基于日志的控制流证明机制 [66], 弥补了软件证明在动态度量方面的不足.

4.3.3 安全代码更新方案

代码更新可以远程修复设备已经披露的漏洞或脆弱性, 还可以远程增加新的特征或系统功能, 开启或禁用某个设备产品的功能, 提升设备生命周期中使用的安全性和灵活性, 最大化节省设备厂商的运维成本.

受限于嵌入式设备资源, 传统代码更新 OTA 中如果代码更新本身存在安全缺陷, 其作为一个基本安全构建块对于物联网环境的整体安全无疑是危险的. 因此, 很多研究者开始考虑设计安全的代码更新方案 (secure code updates). 安全代码更新方案可以归纳为以下 4 类: (1) 基于软件证明的更新. 软件证明可以在设备上提供一个基本的可信执行环境, 可以为安全代码更新提供可信基础, 典型代表是 SCUBA [67]. (2) 基于内存擦除的更新. 这种方案的思路是先采用内存擦除的方法在设备上创建一个 “干净” 的环境, 然后再下载代码更新, 典型代表是 PoSE [68]. (3) 基于安全硬件的更新. 这种方案基于安全芯片来检查设备的固件更新状态, 防止第三方提供错误的固件更新, 通过基于硬件的远程证明机制可以保证设备软硬件配置与状态的完整性, 典型代表是富士通 (Fujitsu) 的基于 TPM 的 ECU 更新方案. (4) 基于免疫代码的更新. 这种方案需要依赖 3 个硬件安全特征, 即不可篡改的免疫代码、安全存储和不可中断执行, 为安全更新协议的设计提供信任锚点. 典型代表是文献 [69].

我们使用 PUF 物理指纹设计了一个实用的轻量级安全代码更新机制 [70], 不依赖于额外的安全芯片和昂贵的硬件安全存储, 将代码更新与 PUF 物理属性绑定, 提升了安全性. 更新前, 通过基于 PUF 的随机内存填充可以确保更新代码安装前没有潜在的恶意代码, 相比基于内存擦除的 PoSE 等方案, 可以节省更多网络带宽; 更新时, 验证代码来源合法且版本最新, 认证待更新设备的身份标识, 防止重放、回滚等攻击; 更新后, 每次加载运行更新代码前都进行完整性验证, 并基于哈希树提高完整性校验的效率.

4.4 可信云

随着云计算在各行各业的广泛应用, 云计算平台自身的安全问题也日益凸显出来. 可信云就是将可信计算和云计算相结合的安全支撑技术, 可以解决云平台在虚拟机信任与监控、可信计算环境构建等重要安全问题. 下面将从虚拟可信平台模块技术、虚拟机自省技术和 Intel SGX 技术 3 个方面对可信云进行阐述.

4.4.1 虚拟可信平台模块 (vTPM)

vTPM (virtual TPM) 是 IBM 公司于 2006 年提出的虚拟化 TPM 方案 [71], 主要应用于虚拟化服务器平台. 虚拟化是云计算环境中的一项重要技术, vTPM 是虚拟化和可信计算相结合的产物. 可信计算与虚拟技术结合, 使得可信技术可以为虚拟环境提供软件安全保障. vTPM 主要解决了虚拟环境如何构建信任根, 如何保障信任根的安全性、可移植性等问题, 为虚拟机提供身份标识、完整性存储、密钥管理、加解密和签名等可信功能.

Open CIT 是一种基于 vTPM 的开放云完整性技术. 通过负载管理、加密和基于硬件的信任链进行控制保证云负载的安全. 该方案使用 Intel TXT 技术, 在系统加载、配置时可以度量服务器固件和软件组件的完整性. 该方案利用 TPM 提供信任根, 并在安全启动时进行使用. Open CIT 的后续项目为 Intel SecL-DC, 其目标是利用硬件信任根建立可信、安全、可控的云环境, 提供具有易用 API 的核心功能库.

我们针对虚拟化平台缺乏信任体系的问题, 提出了基于 vTPM 的可信虚拟化平台两级信任构建系统. 我们采用专用的虚拟机构建了 vTPM 虚拟域, 便于信任根部署和迁移. 在建立信任链时, 我们采用两级信任链构建的方式, 第 1 级是从 TPM 建立硬件到云平台 VMM 的信任链, 第 2 级是从 vTPM 建立 hypervisor 到虚拟机系统的信任链. 在此基础上, 我们针对云计算平台和相关应用的安全需求, 设计并实现了虚拟可信平台模块的架构、功能和接口. 此外, 我们还提出了一个基于虚拟化平台的可信云服务安全管理方案, 采用 vTPM 信任链和可信虚拟机完整性证明技术, 提供了云计算环境虚拟机安全运行状态的可信可视化管理. 该方案解决了云平台的虚拟机安全证明和安全管理问题, 使得远程云管理平台可以安全可靠地验证虚拟机服务器的平台完整性和状态信息.

4.4.2 虚拟机监控技术

在云计算中, 保障虚拟机的安全性具有重要意义. 如果安全软件位于目标宿主机中, 安全软件本身也容易遭到攻击. 于是虚拟机自省 (virtual machine introspection, VMI) 方法被提出, 其将安全软件置于安全虚拟机中, 使用安全虚拟机来监控目标宿主虚拟机. VMI 技术的主要应用包括虚拟机内核完整性监控、入侵检测等. 虚拟机内核完整性监控的代表性方案有 SBCFI [72] 和 SecVisor [73]. SBCFI 通过对目标虚拟机中的函数指针和其指向的代码进行检测, 确保程序的控制流完整性. SecVisor 是一个可以保护商用操作系统内核代码完整性的小型 hypervisor, 该方案确保了只有用户允许的代码可以在内核模式中执行. 基于 VMI 的入侵检测则是通过检测目标虚拟机的状态信息、关键事件等, 判断是否发生入侵行为. 其代表性方案有 Livewire [74], 该方案可以监控特定的主机和网络的变化, 在发生异常事件时可以主动进行干预和控制.

虚拟机自省在可信计算中也有重要应用, 代表性方案有 Terra [75] 和 HIMA [76]. 针对虚拟机监控, 我们提出了使用可信虚拟化技术保障云基础设施安全运行的可信性可视化方案, 利用远程证明的方式将云平台配置、运行软件、网络端口等信息向验证者进行可信性证明, 确保监控管理虚拟机的执行环境和运行状态不被篡改, 普通虚拟机按照正确的配置可信运行.

4.4.3 基于 Intel SGX 的可信云技术

基于 SGX 的机密计算是目前可信云计算的一个研究热点. 微软学者提出了 VC3 [77], 允许用户在云中运行分布式的 MapReduce 运算, 保证结果的正确性和完备性. 伦敦帝国学院 (Imperial College London) 和德莱斯顿大学 (Technical University of Dresden) 的学者提出了 SecureCloud [78], 用于保证大数据应用在云中使用敏感数据时不会遭到数据的泄露. 他们设计和开发了一个分层的架构用于安全创建和部署微服务, 将微服务与大数据应用进行安全集成, 以及在不可信环境中确保这些应用的安全执行. 斯坦福大学 (Stanford University) 的学者提出了 Slalom [79] 方案, 该方案将深度神经网络的应用集成到 TEE 中, 其中一部分运算外包给不可信的硬件并验证结果, 以提高应用的安全性. MesaTEE [80] 给出了一种在公有云中处理安全敏感数据的方法, 具有内存安全、灵活可配置的安全等级、不可绕过的检查等优点. 此外将 TEE 应用到数据库领域也是一个研究热点.

4.5 可信区块链

区块链是当前的研究热点之一, 其公开性和防篡改性使得其在很多场景具有很大的应用潜力. 但是, 区块链自身的安全性也成为人们担忧的一个重要问题, 可使用可信计算技术提升区块链的安全性. 我们将可信计算与区块链的结合称为可信区块链, 下面首先介绍可信区块链研究现状, 然后从可信计算对区块链的安全性增强角度介绍可信区块链主要研究内容.

4.5.1 可信区块链研究现状

区块链大多运行在一个公开透明的环境中, 容易受到恶意攻击. 现有的区块链受限于算法效率等各种因素, 区块链上的交易吞吐量无法满足大部分实际应用场景. 此外, 区块链对于敏感数据的保护也有所欠缺.

将可信计算技术引入到区块链中, 对增强区块链的安全性以及提高共识机制的效率具有积极的作用, 目前的研究进展如下.

Truxen [81] 是一种引入可信计算技术的区块链, 它创新性地提出了一种称为完整性证明 (proof of integrity) 的共识机制. Truxen 使用可信计算和完整性证明来生成区块、执行交易与智能合约、保护敏感数据. PoL (proof of luck) [82] 是一种基于 TEE 的区块链共识机制, 在执行过程中每个节点都会生成一个轮次时间, 轮次时间最短的节点拥有出块权, 通过 TEE 保证节点声称的轮次时间是真实的, TEE 还通过证明机制确保出块结果可以供远程节点验证. Town Crier [83] 是一种经过身份验证的数据馈送方案, 使用 Intel SGX 作为可信执行环境, 结合了区块链前端和可信计算平台后端, 保证智能合约使用的外部数据的可信性. Ekiden [84] 是一种机密、可信的区块链, 可以将智能合约的执行从区块链转移到 TEE 中, 并对其内部数据进行加密. Hyperledger Fabric [85] 是 IBM 研究院提出的一个 Hyperledger 的子项目, Hyperledger Fabric 使用 Intel SGX Enclave 作为 TEE, 在其上执行智能合约, 并维护账本和注册表.

以上可信区块链技术方案中, 可信计算技术改进共识机制或使用 TEE 增强智能合约的安全性, 但相关研究还存在一定的不足. 例如, Truxen 通过引入可信计算节点简化了共识机制的实现, 提升了智能合约的执行效率, 但也会增加网络负载, 同时弱化了区块链的去中心化特性; POL 共识机制也通过引入 TEE 提升了共识协议的效率并保证了共识结果的真实性, 但是在面对恶意攻击节点时安全防护能力不足; Town Crier 专注于获取的外部数据的安全性, 对区块链智能合约执行效率并没有改善, 而且对于以太坊中存在的诸如 “寄生合约” 等问题也缺少解决方案; Ediken 增强了区块链的机密性和可信性, 但其应用平台有一定的局限性; Hyperledger Fabric 对共识算法、加密安全、智能合约都有一定

改进, 但共识算法尚不支持 BFT, 并发控制存在局限性, 整体性能也有待提高.

4.5.2 可信区块链主要研究内容

目前虽然部分可信区块链成熟项目已经可以实现企业级应用落地, 但还有很多共性的关键安全问题有待深入研究. 结合区块链自身安全和可信计算固有优势, 我们将可信区块链的主要研究内容归纳为以下 3 个方面: 一是通过可信执行环境改进共识协议, 通过降低共识协议的复杂性提高区块链的效率; 二是通过可信执行环境保障区块链依赖的计算环境的安全性, 确保智能合约在一个安全可信的环境中执行; 三是通过可信区块链实现分布式网络中大规模物联网设备的节点认证.

5 未来发展趋势及展望

可信计算是一种特有的基于整体安全思想的主动防御技术, 随着网络空间安全技术变革而不断地创新发展. 从早期的 TPM/TCM 安全芯片, 到当前以 ARM TrustZone 为代表的可信执行环境和以 Intel SGX 为代表的 Enclave 架构, 都以 “安全启动、可信执行环境、度量与证明、可信存储” 等可信计算核心关键技术为支撑, 构建可信安全支撑平台, 从而为各种类型终端、边缘端、云端计算环境提供安全基石. 在网络空间安全日益重要的今天, 可信计算无疑是最重要的体系型安全防护技术之一, 拥有广阔的发展和应用前景.

可信计算引领的整体安全架构、主动免疫安全体系, 已经成为网络空间安全技术拼图中不可或缺的一环. 可信计算将以创新理论与技术同计算机体系结构、操作系统安全、可信软件深度融合, 构建更加有效、更加灵活的安全防护体系. 展望未来, 可信计算将在技术和产业的各个方面得到快速发展.

(1) 立足于核心关键技术不受制于人, 我国可信计算必然朝着国产化、自主可控的规模化应用方向发展. 特别是在我国最新发布的《密码法》, 以及已经开始实施的等级保护 2.0, 都强化了对自主可控可信计算技术的使用要求. 随着相关政策和标准的落地, 我国自主可控可信计算有望全面地在国家关键信息基础设施和重要信息系统中得到规模化应用.

(2) 以通用可信执行环境为代表的新型可信计算不断拓展应用领域, 将成为移动互联网、物联网、云计算等领域的主流解决方案. 可信计算将以多种灵活的 TEE 实现方式, 建立模块化、层次化分明的开放性体系架构, 深度融合到智能终端、边缘网关、云服务器等安全平台中.

(3) 可信计算技术和标准体系更加健全, 可信计算测评体系将更加完善. 在政策指导、产业引领、应用驱动下, 可信计算技术和标准将持续提高和完善, 可信计算测评方法、技术和工具也会相应地快速发展. 总之, 完善的可信计算标准和测评体系, 促进我国可信计算产品成熟和技术进步, 带动可信计算产业健康发展.

(4) 可信计算将在信息技术新变革中发挥重要作用, 为人工智能、区块链、物联网、边缘计算等建立重要安全基础. 全球多家巨头企业已经成立机密计算联盟, 将可信计算作为机密计算的重要支撑技术, 以保障人工智能、区块链、物联网等的隐私和安全.

总体来说, 可信计算核心思想已经得到了国内外学术界和产业界的普遍认可, 随着可信计算理论与技术不断完善和发展, 以及相关产业化和工程化经验的不断积累, 可信计算必将在 5G/6G、物联网、人工智能时代迈向新的台阶. 我们应该把握机遇, 以应用技术驱动持续推进可信计算的创新发展, 在网络空间安全领域为国家关键信息基础设施和重要信息系统保驾护航.

致谢 李为、牛海行两位研究生参与了本论文部分内容的写作和讨论, 在此表示衷心的感谢.

参考文献

1 Common Criteria Project Sponsoring Organization. Common Criteria for Information Technology Security Evaluation. ISO/IEC International Standard 15408 Version 2.1. Genevese: Common Criteria Project Sponsoring Organization, 1999. https://www.commoncriteriaportal.org/files/ccfiles/ccpart1v21.pdf

2 Avizienis A, Laprie J C, Randell B, et al. Basic concepts and taxonomy of dependable and secure computing. IEEE Trans Depend Secure Comput, 2004, 1: 11–33

3 Trusted Computing Group. TCG Specification Architecture Overview, Version 1.2. 2003. https://www.trustedcomputinggroup.org

4 State Cryptography Administration. Information security techniques — functionality and interface specification of cryptographic support platform for trusted computing. GB/T 29829-2013 [国家密码管理局. 信息安全技术 — 可信计算密码支撑平台功能与接口规范. GB/T 29829-2013]. http://www.gmbz.org.cn/file/2018-02-06/856e0efa-f970-4173-ae70-b2e933bf40ee.pdf

5 Brickell E, Camenisch J, Chen L Q. Direct anonymous attestation. In: Proceedings of the 11th ACM Conference on Computer and Communications Security, Washington DC, 2004. 132–145

6 Brickell E, Chen L Q, Li J. A new direct anonymous attestation scheme from bilinear maps. In: Proceedings of International Conference on Trusted Computing. Berlin: Springer, 2008. 166–178

7 Chen X F, Feng D G. Direct anonymous attestation for next generation TPM. J Comput, 2008, 3: 43–50

8 Brickell E, Li J. A pairing-based DAA scheme further reducing TPM resources. In: Proceedings of International Conference on Trust and Trustworthy Computing. Berlin: Springer, 2010. 181–195

9 Chen L Q, Page D, Smart N P. On the design and implementation of an efficient DAA scheme. In: Proceedings of International Conference on Smart Card Research and Advanced Applications. Berlin: Springer, 2010. 223–237

10 Brickell E, Chen L Q, Li J. A (corrected) DAA scheme using batch proof and verification. In: Proceedings of International Conference on Trusted Systems. Berlin: Springer, 2011. 304–337

11 Yang K, Zhang Z F, Xi L. Direct anonymous attestation with minimal TPM computational resources. In: Proceedings of China Cryptography Annual Meeting, Zhengzhou, 2014

12 Qin Y, Chu X, Feng D G, et al. DAA protocol analysis and verification. In: Proceedings of International Conference on Trusted Systems. Berlin: Springer, 2011. 338–350

13 Feng D G, Qin Y, Chu X B, et al. Trusted Computing: Principles and Applications. Berlin: Walter de Gruyter GmbH, 2018

14 Zhang Q Y, Feng D G, Zhao S J. Design and formal analysis of TCM Key migration protocols. J Softw, 2015, 26: 2396–2417 [张倩颖, 冯登国, 赵世军. TCM 密钥迁移协议设计及形式化分析. 软件学报, 2015, 26: 2396–2417]

15 Shao J X, Feng D G, Qin Y. Type-based analysis of protected storage in the TPM. In: Proceedings of International Conference on Information and Communications Security. Cham: Springer, 2013. 135–150

16 Zhao S J, Xi L, Zhang Q Y, et al. Security analysis of SM2 key exchange protocol in TPM 2.0. Secur Commun Netw, 2015, 8: 383–395

17 Wang W J, Qin Y, Feng D G. Automated proof for authorization protocols of TPM 2.0 in computational model. In: Proceedings of International Conference on Information Security Practice and Experience. Cham: Springer, 2014. 144–158

18 Shao J X, Qin Y, Feng D G, et al. Formal analysis of enhanced authorization in the TPM 2.0. In: Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security, Singapore, 2015. 273–284

19 Xi L, Yang K, Zhang Z F, et al. DAA-related APIs in TPM 2.0 revisited. In: Proceedings of International Conference on Trust and Trustworthy Computing. Cham: Springer, 2014. 1–18

20 Francois D, Nada E K, Liqun C, et al. First Report on the Security of the TPM. DS-LEIT-779391/D3.2/v1.1. 2019. https://futuretpm.eu/downloads/FutureTPM-D3.2-First-Report-on-the-Security-of-the-TPM-PU-M15.pdf

21 Sun H, Sun K, Wang Y, et al. TrustOTP: transforming smartphones into secure one-time password tokens. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, 2015. 976–988

22 Zhang Y J, Zhao S J, Qin Y, et al. Trusttokenf: a generic security framework for mobile two-factor authentication using trustzone. In: Proceedings of 2015 IEEE Trustcom/BigDataSE/ISPA, Helsinki, 2015. 1: 41–48

23 Liu H, Saroiu S, Wolman A, et al. Software abstractions for trusted sensors. In: Proceedings of the 10th International

Conference on Mobile Systems, Applications, and Services, Low Wood Bay Lake District, 2012. 365–378
24 Ahmad Z, Francis L, Ahmed T, et al. Enhancing the security of mobile applications by using TEE and (U) SIM. In: Proceedings of 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing, Vietri sul Mere, 2013. 575–582
25 Santos N, Raj H, Saroiu S, et al. Using ARM TrustZone to build a trusted language runtime for mobile applications. In: Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems, Salt Lake City, 2014. 67–80
26 Liu D, Cox L P. Veriui: attested login for mobile devices. In: Proceedings of the 15th Workshop on Mobile Computing Systems and Applications, Santa Barbara, 2014. 1–6
27 Li W, Li H, Chen H, et al. Adattester: secure online mobile advertisement attestation using TrustZone. In: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, Florence, 2015. 75–88
28 McCune J M, Parno B J, Perrig A, et al. Flicker: an execution infrastructure for TCB minimization. SIGOPS Oper Syst Rev, 2008, 42: 315–328
29 Winter J. Trusted computing building blocks for embedded linux-based ARM TrustZone platforms. In: Proceedings of the 3rd ACM Workshop on Scalable Trusted Computing. New York: ACM, 2008. 21–30
30 Sierraware. SierraTEE Virtualization for ARM TrustZone and MIPS. 2020. https://www.sierraware.com/open-source-ARM-TrustZone.html
31 Costan V, Lebedev I, Devadas S. Sanctum: minimal hardware extensions for strong software isolation. In: Proceedings of the 25th USENIX Security Symposium, Austin, 2016. 857–874
32 Lee D, Kohlbrenner D, Shinde S, et al. Keystone: an open framework for architecting trusted execution environments. In: Proceedings of the 15th European Conference on Computer Systems. Heraklion: ACM, 2020. 1–16
33 Zhao S J, Zhang Q Y, Qin Y, et al. SecTEE: a software-based approach to secure enclave architecture using TEE. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2019. 1723–1740
34 Sun H, Sun K, Wang Y, et al. TrustICE: hardware-assisted isolated computing environments on mobile devices. In: Proceedings of 2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, Rio de Janeiro, 2015. 367–378
35 Jang J, Choi C, Lee J, et al. PrivateZone: providing a private execution environment using ARM TrustZone. IEEE Trans Depend Secure Comput, 2018, 15: 797–810
36 Zhang N, Sun K, Lou W, et al. Case: cache-assisted secure execution on ARM processors. In: Proceedings of 2016 IEEE Symposium on Security and Privacy, San Jose, 2016. 72–90
37 Yun M H, Zhong L. Ginseng: keeping secrets in registers when you distrust the operating System. In: Proceedings of Network and Distributed System Security Symposium, San Diego, 2019
38 Zhang Y J, Qin Y, Feng D G, et al. An efficient TrustZone-based in-application isolation schema for mobile authenticators. In: Proceedings of International Conference on Security and Privacy in Communication Systems. Cham: Springer, 2017. 585–605
39 Azab A M, Ning P, Shah J, et al. Hypervision across worlds: real-time kernel protection from the ARM TrustZone secure world. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2014. 90–102
40 Ge X, Vijayakumar H, Jaeger T. Sprobes: enforcing kernel code integrity on the TrustZone architecture. 2014. ArXiv: 1410.7747
41 Sun H, Sun K, Wang Y, et al. Trustdump: reliable memory acquisition on smartphones. In: Proceedings of European Symposium on Research in Computer Security. Cham: Springer, 2014. 202–218
42 Brasser F, Kim D, Liebchen C, et al. Regulating smart personal devices in restricted spaces. 2015. https://rucore.libraries.rutgers.edu/rutgers-lib/58513/
43 Fitzek A, Achleitner F, Winter J, et al. The ANDIX research OS—ARM TrustZone meets industrial control systems security. In: Proceedings of 2015 IEEE 13th International Conference on Industrial Informatics, Cambridge, 2015. 88–93
44 Pinto S, Oliveira D, Pereira J, et al. FreeTEE: when real-time and security meet. In: Proceedings of 2015 IEEE 20th

Conference on Emerging Technologies & Factory Automation, Luxembourg, 2015. 1–4
45 Zhang Y J, Feng D G, Qin Y, et al. A TrustZone-based trusted code execution with strong security requirements. J Comput Res Develop, 2015, 52: 2224–2238 [张英骏, 冯登国, 秦宇, 等. 基于 TrustZone 的强安全需求环境下可信代码执行方案. 计算机研究与发展, 2015, 52: 2224–2238]
46 Shor P W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Rev, 1999, 41: 303–332
47 Kassem N E L, Chen L Q, El Bansarkhani R, et al. L-DAA: lattice-based direct anonymous attestation. IACR Cryptol ePrint Arch, 2018, 2018: 401
48 Eldefrawy K, Tsudik G, Francillon A, et al. SMART: secure and minimal architecture for (establishing dynamic) root of trust. In: Proceedings of Network and Distributed System Security Symposium, San Diego, 2012. 1–15
49 Noorman J, Agten P, Daniels W, et al. Sancus: low-cost trustworthy extensible networked devices with a zero-software trusted computing base. In: Proceedings of USENIX Security, Washington, 2013. 479–494
50 Strackx R, Piessens F, Preneel B. Efficient isolation of trusted subsystems in embedded systems. In: Proceedings of International Conference on Security and Privacy in Communication Systems. Berlin: Springer, 2010. 344–361
51 Schulz P K S, Sadeghi A R, Varadharajan V. Trustlite: a security architecture for tiny embedded devices. In: Proceedings of the 9th European Conference on Computer Systems, Amsterdam, 2014. 1–14
52 Brasser F, El Mahjoub B, Sadeghi A R, et al. TyTAN: tiny trust anchor for tiny devices. In: Proceedings of the 52nd Annual Design Automation Conference, San Francisco, 2015. 1–6
53 Zhao S J, Zhang Q Y, Hu G Y, et al. Providing root of trust for ARM TrustZone using on-chip SRAM. In: Proceedings of the 4th International Workshop on Trustworthy Embedded Devices, New York, 2014. 25–36
54 Wurster G, van Oorschot P C, Somayaji A. A generic attack on checksumming-based software tamper resistance. In: Proceedings of Security and Privacy, Oakland, 2005. 127–138
55 Seshadri A, Perrig A, van Doorn L, et al. Swatt: software-based attestation for embedded devices. In: Proceedings of Security and Privacy, Berkeley, 2004. 272–282
56 Li Y, McCune J M, Perrig A. SBAP: software-based attestation for peripherals. In: Trust and Trustworthy Computing. Berlin: Springer, 2010. 16–29
57 Li Y, McCune J M, Perrig A. VIPER: verifying the integrity of PERipherals' firmware. In: Proceedings of the 18th ACM Conference on Computer and Communications Security. New York: ACM, 2011. 3–16
58 Yang Y, Wang X, Zhu S, et al. Distributed software-based attestation for node compromise detection in sensor networks. In: Proceedings of 2007 26th IEEE International Symposium on Reliable Distributed Systems, Beijing, 2007. 219–230
59 Jakobsson M, Johansson K A. Practical and secure software-based attestation. In: Proceedings of Lightweight Security & Privacy: Devices, Protocols and Applications, Istanbul, 2011. 1–9
60 Park T, Shin K G. Soft tamper-proofing via program integrity verification in wireless sensor networks. IEEE Trans Mobile Comput, 2005, 4: 297–309
61 Abera T, Asokan N, Davi L, et al. C-FLAT: control-flow attestation for embedded systems software. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, New York, 2016. 743–754
62 Dessouky G, Zeitouni S, Nyman T, et al. Lo-fat: low-overhead control flow attestation in hardware. In: Proceedings of the 54th Annual Design Automation Conference, New York, 2017. 1–6
63 Asokan N, Brasser F, Ibrahim A, et al. SEDA: scalable embedded device attestation. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2015. 964–975
64 Ambrosin M, Conti M, Ibrahim A, et al. SANA: secure and scalable aggregate network attestation. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2016. 731–742
65 Feng W, Qin Y, Zhao S J, et al. AAoT: lightweight attestation and authentication of low-resource things in IoT and CPS. Comput Netw, 2018, 134: 167–182
66 Liu J B, Yu Q, Liu W, et al. Log-based control flow attestation for embedded devices. In: Proceedings of International Symposium on Cyberspace Safety and Security. Cham: Springer, 2019. 117–132
67 Seshadri A, Luk M, Perrig A, et al. SCUBA: secure code update by attestation in sensor networks. In: Proceedings of the 5th ACM Workshop on Wireless Security, Los Angeles, 2006. 85–94
68 Perito D, Tsudik G. Secure code update for embedded devices via proofs of secure erasure. In: Proceedings of

European Symposium on Research in Computer Security. Berlin: Springer, 2010. 643–662
69 Kohnhauser F, Katzenbeisser S. Secure code updates for mesh networked commodity low-end embedded devices. In: Proceedings of European Symposium on Research in Computer Security. Cham: Springer, 2016. 320–338
70 Feng W, Qin Y, Zhao S J, et al. Secure code updates for smart embedded devices based on PUFs. In: Proceedings of International Conference on Cryptology and Network Security. Cham: Springer, 2017. 325–346
71 Perez R, Sailer R, van Doorn L. vTPM: virtualizing the trusted platform module. In: Proceedings of 15th USENIX Security Symposium, Boston, 2006. 305–320
72 Petroni J N L, Hicks M. Automated detection of persistent kernel control-flow attacks. In: Proceedings of the 14th ACM Conference on Computer and Communications Security. New York: ACM, 2007. 103–115
73 Seshadri A, Luk M, Qu N, et al. SecVisor: a tiny hypervisor to provide lifetime kernel code integrity for commodity OSes. In: Proceedings of 21st ACM SIGOPS Symposium on Operating Systems Principles, Stevenson, Washington, 2007. 335–350
74 Garfinkel T, Rosenblum M. A virtual machine introspection based architecture for intrusion detection. In: Proceedings of Network and Distributed System Security Symposium, San Diego, 2003. 191–206
75 Garfinkel T, Pfaff B, Chow J, et al. Terra: a virtual machine-based platform for trusted computing. In: Proceedings of the 19th ACM Symposium on Operating Systems Principles, Bolton Landing, 2003. 193–206
76 Azab A M, Ning P, Sezer E C, et al. HIMA: a hypervisor-based integrity measurement agent. In: Proceedings of 2009 Annual Computer Security Applications Conference, Honolulu, 2009. 461–470
77 Schuster F, Costa M, Fournet C, et al. VC3: trustworthy data analytics in the cloud using SGX. In: Proceedings of 2015 IEEE Symposium on Security and Privacy, San Jose, 2015. 38–54
78 Kelbert F, Gregor F, Pires R, et al. SecureCloud: secure big data processing in untrusted clouds. In: Proceedings of the Conference on Design, Automation & Test in Europe, Lausanne, 2017. 282–285
79 Tramer F, Boneh D. Slalom: fast, verifiable and private execution of neural networks in trusted hardware. https://arxiv.org/abs/1806.03287
80 The MesaTEE Team. MesaTEE: a framework for universal secure computing. 2020. https://mesatee.org/
81 Zhang C. Truxen: a trusted computing enhanced blockchain. 2020. https://arxiv.org/abs/1904.08335
82 Milutinovic M, He W, Wu H, et al. Proof of luck: an efficient blockchain consensus protocol. In: Proceedings of the 1st Workshop on System Software for Trusted Execution, Trento, 2016. 1–6
83 Zhang F, Cecchetti E, Croman K, et al. Town Crier: an authenticated data feed for smart contracts. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, 2016. 270–282
84 Cheng R, Zhang F, Kos J, et al. Ekiden: a platform for confidentiality-preserving, trustworthy, and performant smart contracts. In: Proceedings of 2019 IEEE European Symposium on Security and Privacy, Stockholm, 2019. 185–200
85 Brandenburger M, Cachin C, Kapitza R, et al. Blockchain and trusted computing: problems, pitfalls, and a solution for hyperledger fabric. 2020. https://arxiv.org/abs/1805.08541

Trusted computing theory and technology in innovation-driven development

Dengguo FENG[1,2], Jingbin LIU[2,3], Yu QIN[2*] & Wei FENG[2]

1. *State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing* 100190*, China*;
2. *Trusted Computing and Information Assurance Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing* 100190*, China*;
3. *University of Chinese Academy of Sciences, Beijing* 100049*, China*
* Corresponding author. E-mail: qin_yu@tca.iscas.ac.cn

Abstract Trusted computing is based on a hardware security mechanism establishing a trusted computing environment and comprehensively enhances the system and network trust from the architectural perspective. With the development of information technology and continuous emergence of new application scenarios, security threats in the cyberspace are becoming increasingly serious; hence, trusted computing is actively researched in both academia and industry to find solutions against such treats. This paper summarizes the development process of trusted computing theory from the perspective of innovation and development. The study centers around one of the author's research results in trusted computing over the past 20 years. It proposes a trusted computing technology architecture that covers two method foundations, three trust cores, and four key technologies. Furthermore, the paper summarizes important research problems in mobile trusted computing, quantum-resistant trusted computing, trusted Internet of Things (IoT), trusted cloud, and trusted blockchain, elaborating on the integration and development of trusted computing in these fields. In mobile trusted computing, the design and implementation of a trusted execution environment architecture with software/hardware co-design is the focus of research. Another two important research issues in mobile trusted computing are the runtime security isolation and protection of the mobile operating system's kernel and trusted execution environment-based mobile application security protection. Due to the characteristics of embedded environments and limitation of resources, the construction of lightweight trusted roots, efficient and secure software attestation, practical secure code update mechanism, and swarm device attestation are important issues for further research in trusted IoT. In new scenarios such as quantum-resistant trusted computing, trusted cloud, and trusted blockchain, trusted computing is also constantly expanding its application boundaries and playing an increasingly important role. Finally, this paper looks ahead and discusses the development trends in trusted computing.

Keywords trusted computing, trusted execution environment, mobile trusted computing, quantum-resistant trusted computing, trusted Internet of Things, trusted cloud, trusted blockchain

Dengguo FENG was born in 1965. He is a professor and a Ph.D. supervisor. His current research interests include network security and information security.

Jingbin LIU was born in 1992. He is a Ph.D. student. His current research interests include trusted computing and information security.

Yu QIN was born in 1979. He is a Ph.D. in computer science and senior engineer. His current research interests include applied cryptography, system security and trusted computing.

Wei FENG was born in 1986. He is an associate researcher. His current research interests include system security and trusted computing.