基于社会学信任理论的软件可信性概念模型

杨 曦,罗 平,贾古丽
(清华大学软件学院信息系统安全教育部重点实验室,北京 100084)

摘 要: 可信性作为软件的一种复杂的高复合概念,几十年间都未能取得实质性进展和突破. 本文在对可信性权威定义分析的基础上,论证了这些定义所涉范围彼此矛盾且不相容,进一步说明从本质出发研究软件可信性概念模型的重要性和必然性. "可信"一词源于社会学,所以应该从社会学的信任理论出发来探讨软件可信性的本质. 本文在上百篇经典社会学信任理论文献上构建出信任体系模型 STM,并与软件的信任体系进行了对比和映射,提出基于社会学信任理论的软件可信性概念模型 STCM. 在 STM 和 STCM 的基础上给出软件可信性概念模型的定义系统. 最后通过度量评估实验验证了模型是可行的、有效的,为软件可信性的发展提供了新的研究方向.

关键词: 软件可信性; 信任理论; STM; STCM; 可靠性; 安全性
中图分类号: TP311 **文献标识码:** A **文章编号:** 0372-2112 (2019)11-2344-10
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.11.016

The Concept Model of Software Trustworthiness Based on Trust-Theory of Sociology

YANG Xi, LUO Ping, GUL Jabeen
(*The Key Laboratory for Information System Security, Software School, Tsinghua University, Beijing* 100084, *China*)

Abstract: As a complex high-composite software concept, the trustworthiness research has failed to make substantial progress and breakthroughs during these decades. After analyzing the authority definitions of trustworthiness, this paper demonstrates that the scope of them is contradictory and incompatible. It further illustrates that the research of software trustworthiness conceptual model from its essence is very important and necessary. The term "trustworthiness" originates from sociology; so we should discuss the essence of software trustworthiness based on trust-theory of sociology. This paper constructs a trust system model STM based on hundreds of classical sociological literature on trust-theory. After comparing and mapping with software trust system, the paper proposes a software trustworthiness conceptual model STCM based on trust theory of sociology. Based on STM and STCM, a complete concept definition system of software trustworthiness is presented. Finally, STCM is proved to be feasible and effective by the measurement and evaluation experiment, which provides a new research direction for the advancement of software trustworthiness.

Key words: software trustworthiness; trust-theory; STM; STCM; reliability; security

1 引言

可信软件及软件可信性研究是信息技术高速发展、人们不得不面临的新兴课题和全新挑战[1]. 而信任问题几乎在任何领域、但凡可能存在不确定性或不可预知结果的任何情况下,都至关重要.

然而,从 1985 年美国国防部制定 DoD5200.28-STD (即可信计算机系统评测标准 TCSEC)正式提出可信计算(Trusted Computing)概念以来[2],作为一种复杂的高复合概念[1],可信性几十年间在概念认知上却从未有过共识,也没有取得更多实质性进展和突破[3]. 主要体现在:(1)未给出统一、准确的定义;(2)很多学者仅把传统的软件质量、软件工程、软件过程等,冠以软件"可信"研究,换汤不换药;(3)专业术语得不到统一[3],对后序研究形成巨大认知屏障;(4)不同领域、不同方向的研究人员对软件可信性的认识各执己见、自圆其说. 综上,学术界对"软件可信性"概念的本质和源义认识不清导致的诸多问题,使得软件可信性概念及其定义

收稿日期:2018-09-11;修回日期:2019-03-25;责任编辑:李勇锋
基金项目:国家自然科学基金重点项目(No. 90818021);核高基重大专项(No. 2012zx01039-004-46);国家发改委信息安全专项(No. 2012-1424)

的明确、统一已迫在眉睫.

2　研究背景

人们在认识软件可信性的“本质”上从一开始就出现很多问题. 有些研究将可信软件与可信计算、可信系统、甚至可信网络混为一谈,且用到的术语都没统一,先后出现:Trustworthiness, Dependability, Confidence, Assurance 等等.

概念定义工作与理论研究工作、实证检验工作一样重要. 概念模型可以帮助研究人员将科学与现实世界进行正确对接,有了准确的概念基础,科学研究成果才具有更重大的实用价值[4]. 科学逻辑实证主义的观点认为,新式理论必须概念在前,而后才具有可操作性,而实证检验的结果才能进一步用于验证概念理论[5].

为进一步论证以上观点,课题组把当前被广泛接受和认可的、权威学术科研机构提出的软件“可信性”的定义逐一罗列出来(如表 1 所示),并根据软件质量属性特征和模型[6]绘制出这些定义所涉及的软件属性特征的语义范围(如图 1 所示).

从图 1 的比较与分析中,我们发现以下一些问题:(1)这些定义所囊括的语义范围彼此既不完全相容,也不完全相斥,相互矛盾且不可统一;(2)这些定义几乎都没有可令人信服的依据和来源,更无有效性验证,带有领域偏颇性.

那么,什么是可信?“可信”的本质是什么?众所周知,“可信”一词源于社会学,那是否可以从它的根源——社会学开始讨论“可信”本质?

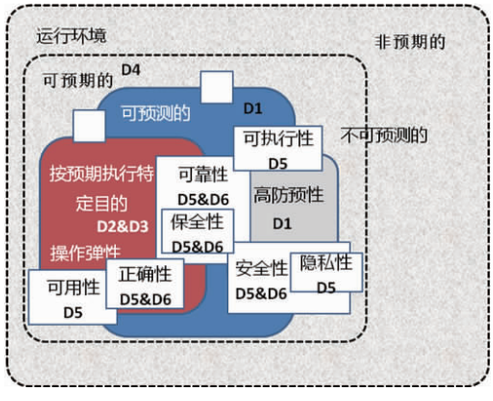

图1　国际权威软件可信性定义涉及范围比较

3　社会学信任体系模型

信任理论研究历史已近百年. 一般认为,“信任”通常被定义为一个人依据目标信任体的特征判断是否信赖它的意愿程度[7]. 本文研究重点放在“目标信任体的特征”,即软件的可信特征和属性,通常用术语“trustworthiness”[8]或“trusting beliefs”[9,10]表示,结合软件领域现有研究的惯用语,我们选择“trustworthiness”作为软件“可信性”的统一术语.

对于“信任”更完整的理解应该进一步考虑信任关系(relationship)的动态演化[11,12]. 社会学认为信任的动态演化有几个阶段[10~13]. (1)初始型信任:这种信任意味着信任方在延伸信任前会对形势做出理性决定[13]. 初始信任的一种形式是基于演算的信任[12],信任方通过评估成本和效益来扩展信任. 初始型信任由于缺乏对目标信任体的特征认知,更多基于自身的认知判断,而这就与信任方的自身特征(trustor's characteristics,如图 2 所示)息息相关了. (2)认知型信任:这种信任意味着信任方对目标信任体有了充分了解后,可以在某些情境下预测目标信任体的行为,所以,认知型信任是基于对目标信任体特征(trustee's characteristics,如图 2 所示)认知后的判断行为. 图 2 是基于社会学的信任体系模型(Sociology Trust-System Model, STM),该模型是建立在 2.1 至 2.5 节的社会学信任理论的基础之上,下面将分别阐述.

3.1　信任方的特征

“信任”首先是由信任方发起的意向行为,而这种信任显然是基于对目标信任体可信性(trustworthiness)的认知. 而信任方自身的特质也是影响一方对另一方信任的因素之一,社会学信任理论权威文献[8,10,11,14]认为,信任方特质包括“信任倾向”(Propensity to Trust)和“制度信任”(Institution-based Trust),如图 2 所示.

信任倾向. 指一个人一贯倾向于在大部分情境和人群里愿意依赖别人的程度,通常这是信任方选择相信的内在因子. 信任倾向可能是在陌生环境里最相关的“信任前件”(trust antecedent)[8]. 这种“构造”(Construct)主要来源于人的性格或心理特质[9]. 其有两个子构造,人道信仰和信任立场.

制度信任. 指人们相信“有利条件”更加有助于人类通过努力而取得成功. 例如,在互联网环境下,“有利条件”意味着法律、监管、业务和技术环境等可以支持成功的诸多元素. 制度信任也有两个子构造,结构保障和情境常态.

软件(目标信任体)的使用者(用户)往往是有不同特质(需求)的人群. 从信任倾向各异的集群中提取信任方特征的共性,这个过程就是对用户软件可信性需求的获取和提炼. 因此,本文将信任倾向映射为“软件可信性需求规约”(如图 3 的 STR, Software Trustworthiness Requirements),而“制度信任”会对用户的心理预期造成影响,最终也会反映到 STR 上,而 STR 则会影响软件可信性的研发.

表 1 软件可信性国际权威定义

编号	年份	术语	定义或观点	所覆盖软件属性	出处
D1	2012	Trusted	所谓的可信组件、可信操作或可信过程指的是,在几乎任何运行情况下其行为都是可预测的,且对于所有意图入侵的软件、病毒及一定级别的物理干扰都具有高防御性。		ISO/IEC15408
D2	2006	Trustworthiness	一个系统按预期执行特定的目的,在需要的时候,具有一定的操作弹性,并且不会带来非预期的副作用、行为或可利用的漏洞.		CNSS IA4009
D3	2007	Trustworthy	一个如果总是按照其设定的预期目标执行的实体才是值得信任的.		TCG
D4	1998	Trustworthiness	可信性是一个系统值得信任的保证,它能确保在意外发生时都能按预期执行,比如环境中断、人为操作失误、恶意攻击及设计和实现上的错误。可信系统应该能强化信任,让人类相信其能持续产生预期行为,并且不容易被侵覆.	正确性、可靠性、安全性、隐私性、保全性、可生存性等	NRC
D5	2006	Trustworthiness	软件系统的可信性由以下特征决定:正确性、保全性、可用性、可靠性、可执行性、安全性、隐私性.	正确性、保全性、可用性、可靠性、可执行性、安全性、隐私性	TrustSoft
D6	2010	Trustworthiness	可信性可以分解为一系列与可信相关的属性集合,包括保全性、可靠性、安全性、正确性、可用性及其他相关属性[8],其中安全性是软件可信性的重要组成因素,且以安全性为核心的软件可信性必须基于集成化过程中持续设计、规约、确保.	保全性、可靠性、安全性、正确性、有用性等	NIST

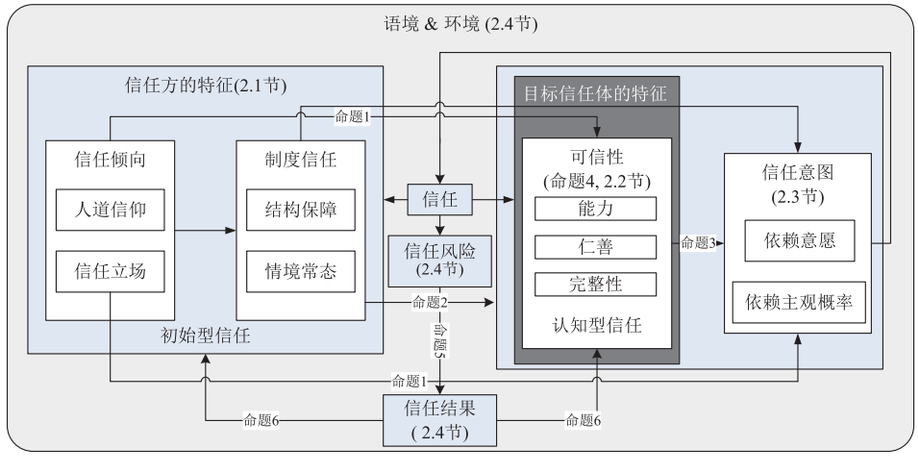

图2 社会学信任体系模型

另一方面,虽然文献[8,10]将制度信任归为信任方的特征,但我们注意到,制度信任是受外部环境、社会进步或技术发展所影响的(比如,法律、法规、标准、合约、规则等),良好的外部环境更容易促使信任方达成信任. 而对于可信性而言,与之相关的软件知识产权、开发标准、行业基本规范或准则等(比如,盗用他人源代码、侵犯知识产权的软件值得相信吗?),都会对软件的信任体系产生影响. 由此,本文将信任方的“制度信任”映射为软件可信模型的外部条件“软件可信领域的法规及标准”(如图 3 所示),其对软件的可信性将起到指导和影响作用,也是软件运行的大环境因子.

信任方特征到软件可信概念模型的映射关系如表 2 所示.

表 2 信任方特征的映射

社会学信任构造		映射	软件可信概念模型
信任倾向	人道信仰	→	软件可信性需求规约(STR)
	信任立场		
制度信任	结构保障	→	STR 及软件可信领域的法规及标准
	情境常态		

3.2　目标信任体的特征:可信性的概念

要想更好地理解,为什么给定的一方对另一方会有或多、或少的信任量,最重要的便是考虑目标信任体的特征[8]. 文献[8~10]对大量社会学文献进行统计分析,得出结论:能力(ability)、仁善(benevolence)、完整性(integrity)和可预测性(predictability)覆盖了91.8%的基于目标信任体特征的人际信任定义,这四个特征已几乎涵盖了目标信任体"可信性"的主要部分."可信"作为一种模糊理论,讨论清楚90%以上的概念范围就足以理解可信的本质. 然而,软件有别于社会信任理论,有其自身的特征,所以我们必须区别对待.

能力:能力已成为被讨论得最为广泛、最为常见的可信元素之一[7~10,14]. 能力是一组在某些特定领域内促使其"附体"具有可信影响力的技能、能力和特征集合,并需具备在风险情境下应有的通常智慧. 能力是有领域性的. 我们把能力可信(Ability Trustworthiness)记作一个三元组的特征函数:

$$\mathrm{AT}(K,D,R)$$

$K\in\{k\mid k$ 是完成特定任务所需的知识和技能$\}$

$D\in\{d\mid d$ 指的是能力具有领域相关性$\}$

$R\in\{r\mid r$ 指的是风险情境下的解决能力$\}$

仁善:仁善指的是目标信任体被相信总是能对信任方做"好"事的程度,而不考虑个人的利益动机[8]. 其同义词包括:忠诚、开放、关爱或支持等. 我们将仁善可信(Benevolence Trustworthiness)记作一个三元组特征函数:

$$\mathrm{BT}(P,C,Y)$$

$P\in\{p\mid p$ 指的是正面、积极的影响$\}$

$C\in\{c\mid c$ 指有忠诚、开放、关怀或支持等特征$\}$

$Y\in\{y\mid y$ 指的是仁善与领域无关$\}$

完整性:完整性被定义为目标信任体被相信具有信任方认可的好的品德和伦理原则的程度[7~10,15]. 其同义词包括公平、公正、一致性和履行承诺等. 因而,完整性是一种价值中立的特征.

可预测性:可预测性意味着一方相信他方的行为(不管是好的还是坏的)都具有足够的一致性特征,从而可以预知在特定情况下的他方行为.

完整性与可预测性是相辅相成的,完整性包含一致性特征,而一致性最终保障了可预测性,文献[8,15]也是将可预测性合并到完整性,于是这两种特性合并定义为完整可信性(Integrity Trustworthiness,IT),IT为三元组,记作:

$$\mathrm{IT}(S,N,Y)$$

$S\in\{s\mid s$ 指的是与公平、公正、一致性、履行承诺及可预测性相关的原则$\}$

$N\in\{n\mid n$ 指的是价值中立的特征$\}$

$Y\in\{y\mid y$ 指的是仁善与领域无关$\}$

目标信任体特征是本文的研究重点,这部分内容如何映射较为复杂,将单独在3.1节结合软件可信特征专门分析.

3.3　信任意图

信任意图指的是一方愿意依赖,或打算依赖他方的意愿程度. 信任意图基于人类自身特征和主观判断,甚至是一种心态情绪上的变化函数. 而软件的使用者也是各类人群,每个人的主观意识和感觉是多变、矛盾、不可靠的,导致对软件可信的认识可能是非协调信息,按照经典逻辑平凡推理的观点,从非协调的(矛盾的)前提可以推导出任何结论,这就意味着含有非协调信息的系统刻画、数据是毫无意义的. 这样的非协调信息必须寻求第三方专业机构来解决统一问题. 因此,如图3所示,我们将"信任意图"映射为权威评估机构(Authoritative Evaluation Agency,AEA). 然而,这并非说明信任方的信任意图在软件领域就不起作用了,相反,软件领域中信任方的信任意图事实上会间接转化为软件用户对软件可信性的进一步需求(STR),从而最终体现到权威评估机构对软件可信性的评估报告中. 信任意图的映射结论如表3所示.

表3　信任意图的映射

社会学信任构造		映射	软件可信概念模型
信任意图	依赖意愿	→	权威评估机构(AEA)
	依赖主观概率		

3.4　风险、动态演化与语境条件

3.4.1　人际关系风险与软件可信风险

往往有风险才会引发"信任"需求. 依赖他方就要求信任方必须冒某种风险,即目标信任体可能会有意无意地不履行预期责任的风险. 信任程度则会影响信任方愿意承担的风险级别.

诚如没有百分百值得相信的人类,也没有百分百正确、可靠、完全值得信任的软件,如何正确地预测、评估、度量可信风险成为了研究关键. 而信任风险因信任而生,是否选择信任追根溯源还是来自于对信任方和目标信任体特征函数的认知程度.

3.4.2　信任的动态演化

对于"信任"更加完备的理解必须考虑动态演化问题. 信任级别会随着参与双方的不断互动而逐渐演变甚至进化[8,10,11,14]. 信任的动态演化本质体现在图2的从风险"后果"到已认知的目标信任体的可信特征及信任方的可信特征的反馈环上. 信任行为的后果(有利的或不利的)将会通过目标信任体与信任方的下一次互动而表现出的特征,从而再次间接影响信任.

3.4.3 信任的语境条件

尽管信任级别可能是一个常量,然而信任的具体后果则由语境因素决定[8,10]. 同样,信任前件(antecedent)评估也会受所处语境影响. 总之,信任方对关系语境的感知和演绎会影响对信任的需求及对可信性的评估. 要想对信任有清晰认识就必须理解语境是如何影响可信性认知的. 所以,如图 2 所示,信任体系的各个因素都会受到所处环境影响.

类似地,软件与用户之间的交互显然也会受到外界环境的影响,软件是否可信与其运行环境、上下文关系或相关外部元素(包括硬件、网络、外设、关联系统、接口、用户、数据库、行业标准与法规等等)紧密相关,这些就是我们所说的软件语境. 所以,如图 3 所示,软件可信性模型受环境变量 $E(t,r)$ 约束和限制,这里的 t 指的是时间,因为环境随时间变化而不同的,r 指的是可信性模型中各个元素之间及各元素与环境之间的关系(relationship).

关于风险、动态演化及语境到软件可信概念模型的映射,如表 4 所示.

表 4 风险、动态演化及语境的映射

社会学信任构造	映射	软件可信概念模型
人际关系风险	→	软件可信风险损失
人际信任动态演化	→	软件可信性的动态演化
人际信任语境	→	软件环境变量 $E(t,r)$

4 软件可信性概念模型

前述章节已对 STM 进行了详细阐述、分析,部分已进一步映射到图 3 的软件可信性概念模型,其中最核心部分“软件可信性”如何从 STM 映射而来,单独在本节进行阐述.

4.1 软件的特征:软件可信性的概念

软件信任体系中,软件可信性是核心内容. 2.2 节指出,作为一个集合,能力、仁善和完整性这三个特征变量已几乎涵盖了目标信任体“可信性”的主要部分(超过 91.8%). 并且这三个特征变量对信任的形成都具有各自有意义的、独特的关系,且信任是信任方感知到目标信任体的能力、仁善、完整性的特征函数. 所以,对三个信任目标体特征我们分别进行阐述和分析.

4.1.1 能力可信(AT)的映射

如前所述,能力可信是一个三元组,记作:AT(K,D,R). 然而,软件有别于人类,从 STM 推演出软件可信属性时,我们必须结合软件的特征加以考虑. ①领域相关性 D:在 ISO/IEC9126 标准中,将软件的质量属性分为功能性(Functionality)、可靠性(reliability[16,17])、可用性(usability)、效率(efficiency)、maintainability(可维护性)、protability(可移植性)等六个复合性质,而这六个复合性质又进一步分解出更多子特性[16](如表 5 所示). 根据能力可信的三元组特征,软件的功能性是领域相关的,辅助人们完成某些业务领域的特定任务. 然而,功能性是软件系统的一个复合性质,又进一步分解细化为一组子特性或子属性(如表 5 所示). 根据 ISO 给出的定义,适宜性倾向于软件设计者和客户的能力,也与双方的沟通、合作有很大关系,是对软件需求、设计过程的要求,而非软件本身的能力[17]. 互操作性是正确性或适宜性的子集,因为交互的目的最终也是为了获得正确的结果或合适的功能;功能合规性指的是软件产品符合标准、公约、法律法规及各类约束的能力,这一点对于所有软件属性都是必要的[15,16],为此我们将其映射为软件的基本准则可信(Basic Standard Trustworthiness),如图 3 所示. 基本准则可信的定义后续再述. ②风险情境 R:能力可信应该具有灵活地防范风险、抵御风险的技能,而软件的一切风险都源自于错误(fault)[18],错误导致的软件失效才会造成最终的风险,正确性和安全性的定义都是建立在术语“错误”之上的[15~18],而适宜性和互操作性更倾向于用户体验与交互方面的特征,属于软件的外部特征,不具备“因软件内部错误失效而导致风险”这个特点,也进一步证明这两个属性可以排除. ③知识技能性 K: K 指的是完成特定任务所需的知识和技能,对于软件而言,所有软件属性的达成都需要软件行为及相应数据的支撑,这其实就是各属性所具备的对应技能,因此知能技能 K 是任何属性都需要具备的.

除功能性之外,软件其他属性是否具备 AT 的三元组特征呢? 从 ISO/IEC 9126 对六个软件复合性质的定义,可靠性也具备 AT 三元组特征(如表 5 所示),而可用性、效率、可维护性、可移植性等并不具备,可用性强调的是软件是否容易理解、学习、使用等,属于软件的外部特性(就好比在人类社会里,不能因为对方长得丑陋、肢体残缺,就认为不值得相信,是一样的道理),而效率、可维护性、可移植性等更不在软件可信讨论范畴内. 因篇幅所限,这几个复合性质的定义及子属性的分析,本文不做展开,读者可自行查阅文献[15~18].

综上,从 STM 的三元组 AT(K,D,R)可以映射出软件的能力可信包含正确性、安全性和可靠性,从 2.2 节的“制度信任”及 ISO/IEC 9126 标准的各个软件性质都需具备的合规性我们都可以得到,软件的可信性包含“基本准则可信”,此处的基本准则指的是软件产品必须要符合软件行业的标准、公约、法律法规及各种约束等.

4.1.2 仁善可信(BT)的映射

如前所述,BT 是一个三元组:BT(P,C,Y). 软件如

何产生正面、积极的影响(元组 P)? 又要有忠诚、开放、关怀或支持的特征(元组 C),且与领域无关(元组 Y). 这些都是人类内在特质、性格等的外在表现,那软件的内在特质是什么? 是什么让软件对外表现出如此丰富多彩的特性? 软件的根源、本质就是源代码(源代码为软件的真实身份),即我们通常所说的"指令". 指令是人类意志的外在反映,也是软件由内而外表现出来的"性格"和"特质". 所以,根据 BT 三元组特征,从软件源码的角度去讨论这部分的可信性质,是正确的方向. 既然软件必须能产生正面、积极的影响,而且软件要对用户忠诚、开放、关怀或支持,那么,其(身份)代码必须来源清楚,不具欺瞒性(盗用他人代码、侵权、大量使用开源代码、黑客代码等),这就是我们首次提出的"身份可信",也与 STM 提出的仁善可信本意一致. 当前热门的实名计算研究[19]也再一次论证了我们的观点,文献[19]指出,软件的隐患绝大部分是由于软件自身问题导致的,因为软件开发者忽视安全、故意植入、恶意潜伏等行为直接或间接造成各种隐患,软件已不仅仅是诞生之初的辅助工具,更攸关整个社会生态体系的重要环节,从而,软件的身份认定就显得十分必要. 而且,软件身份可信也确实是领域无关的,对所有的软件都有同样的要求,故而身份可信满足 BT 三元组的特征. 从反证法的角度,我们也无法举出一个反例可以驳倒这个结论. 因此,我们将 BT 映射为 Identity Trustworthiness(身份可信),如图 3 所示.

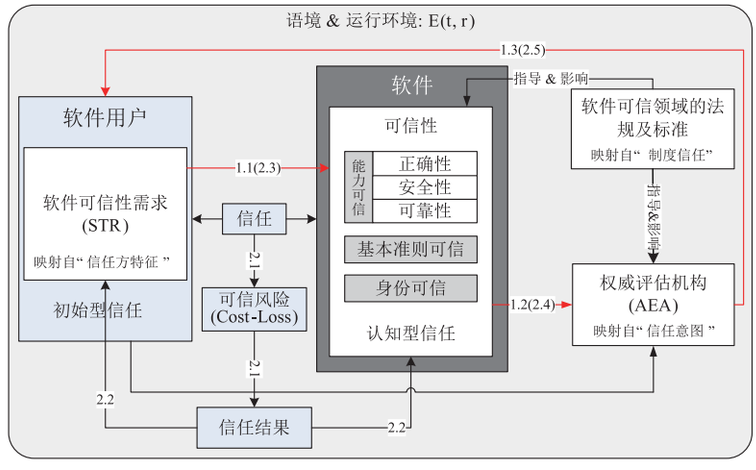

图3　软件可信性概念模型

表 5　软件质量属性与 AT 的映射关系

特征	子特征	定义	K	D	R
功能性	适宜性	软件产品为特定任务和用户目的提供适合的功能集的能力.	√	√	
	正确性	软件产品在可接受的精确度范围内提供正确或一致结果或效果的能力.	√	√	√
	互操作性	软件产品与一个或多个指定系统交互的能力.	√	√	
	安全性	软件产品保护信息和数据不被未经授权的人员或系统读取或修改的能力,并且经过授权的人员或系统不被拒绝访问的能力.	√	√	√
	功能合规性	软件产品在功能方面坚持标准、公约或法律法规及类似约束的能力.	√	√	√
可靠性	成熟性	软件产品避免失效(软件错误导致的结果)的能力.	√	√	√
	容错性	软件产品在软件错误或侵犯指定接口的情况下,维持指定性能水平的能力.	√	√	√
	可恢复性	软件产品重建指定性能水平及恢复数据(因失效直接影响的)的能力.	√	√	√
	可靠合规性	软件产品在可靠性方面坚持标准、公约或法律法规及类似约束的能力.	√	√	√

4.1.3　完整可信(IT)的映射

IT 也是一个三元组,记为 IT(S,N,Y). IT 有一个非常重要的特征,即必须是价值中立的. 于软件而言,所谓价值中立指的是,不管软件是正常运行,还是发生故障,都必须保证最终结果的一致性. 这事实上就是软件的属性——完整性[15,18]的特征,而完整性因前后要保持一致性,故也具备可预测的特征且与领域无关.

然而,文献[18]阐述证明了软件的完整性是安全性的子集. 安全性做为一种公共概念,它并不是一种原子属性,而是一种复杂的的复合属性[15,18]. 所以,IT ⊆

AT,可以将 IT 排除出软件的信任体系,因为其内容已经包含在 AT 中.

表 6 目标信任体特征的映射

人文社会学信任理论	映射	软件可信概念模型
能力可信(AT)	→	正确性、安全性、可靠性
		基本准则可信
仁善可信(BT)	→	身份可信
完整可信(IT)	→	安全性

由此,我们得到 STCM 目标信任体到软件特征的映射关系,如表 6 所示. 通过 3.1 节的分析,可以得到如图 3 所示的软件可信性概念模型,其中的软件可信性包括能力可信(含正确性、安全性和可靠性)、基本准则可信和身份可信,这三个方面都是从 STM 推导而来,具有社会学信任问题的本质特征.

4.2 软件动态演化模型

如图 3 所示,软件动态演化性体现在两方面. ①软件可信需求 STR 对于最终软件的可信性有指导和影响(箭头 1.1 所指示). 正式发布的软件产品(或测试阶段的软件)在运行和测试过程中形成的记录和证据提供给第三方权威评估机构(AEA),通过专业工具和方法进行可信性评测(箭头 1.2 所指示),AEA 给出最终的可信(风险)评估报告,提交给用户参照 STR 进行审核(箭头 1.3 所指示),如果用户根据评测报告给出“不信任”评级,则必定会重新修定 STR,或者开发团队根据评测报告修改、增强软件可信性,从而再次启动上述过程,如果选择“信任”软件(即,软件可信风险在用户可接受的范围),那么将会启动第②种动态演化情境. ②如前所述,用户一旦选择“信任”软件而正式投入使用,就进入了软件风险承担(箭头 2.1 所指示),软件运行使用过程中,非可信因素迟早会引发风险(或多或少),造成一定的风险“后果”,而后果造成的损失会积累起更多的记录和证据供软件可信性进一步评估使用(如图 3 的 2.2,2.3,2.4,2.5 的动态演化过程).

5 软件可信性概念定义体系

可信性是软件信任体系的核心,但必须把它放在整个软件信任体系中去分析. 另一方面,为了软件可信性评估度量结果有统一的量纲,必须对软件可信性进行全面的、系统的定义.

根据国际标准 IEEE Std 610-12-1990 对缺陷、故障、失效的定义,我们有如下定义.

定义 1 缺陷(Fault):计算机程序中一个不正确的步骤、过程或数据定义称为软件的缺陷.

定义 2 失效(Failure):系统或组件无法在指定的性能要求内执行所需功能的行为,称为软件的失效.

定义 3 功能障碍(Disfunction):当遇到故障时,系统无法执行所需功能的行为,称为软件的功能障碍.

注意上述失效与功能障碍是有区别的,如果在软件开发中具有容错设计,则软件可能存在失效而无功能故障,在无容错设计的软件中失效就变成了功能障碍. 定义 1 并未考虑到当前软件(尤其是互联网环境下)的新现象——信息安全问题,为了补充和强调软件安全问题,我们对定义 1 进行了扩展.

定义 4 缺陷(Fault'):计算机程序中一个不正确的步骤、过程、数据定义或违背系统安全原则、策略,引发系统安全问题的设计、编码及数据定义统称为软件的缺陷.

定义 4 与 IEEE Std 610-12-1990 传统缺陷定义兼容,同时考虑并强调了软件的安全问题,故本文提及的缺陷指的是定义 4. 在定义 4 扩展软件缺陷定义的基础上,下面我们给出软件漏洞的定义.

定义 5 漏洞(Vulnerability):如果软件中的缺陷通常情况下不会引起系统失效(不影响软件的正常功能),但有可能被攻击者成功利用后引发软件去执行额外的恶意代码(做超出设计范围的事情),或导致信息的泄漏,则称该缺陷为漏洞,称这些现象为入侵. 如软件中的缓冲区溢出漏洞、SQL 注入漏洞、软件后门漏洞、中间人攻击漏洞、默认命令漏洞、加密脆弱性漏洞等.

注意:上面给出的软件缺陷和漏洞的定义跟传统定义不一样. 这里的缺陷包括软件开发规格说明书中不正确的需求、设计等. 缺陷是静态的,漏洞是动态的,漏洞由缺陷引起,是缺陷的一部分,但缺陷不一定是漏洞. 我们把由软件引起的系统故障、不能按指定性能完成的功能、执行额外恶意代码、泄漏信息等问题都归结为由缺陷和漏洞引起.

根据对人类社会可信性的本质性质的分析,我们给出软件可信性的一般定义.

定义 6 软件可信性:指在规定的环境 $E(t,r)$ 下及规定的时间 t 内,即使在其它因素(如对软件的不当操作)的干扰或人为的攻击下,软件的身份、基本规范或准则、能力这三方面证据赢得信任评判方相信的程度,并相应称为身份可信、基本规范或准则(即行业基本规则)可信和能力可信.

注:身份指的是软件源代码来源清晰、真实、可靠;基本规范或准则可信指的是软件设计、需求符合该行业基本规则或符合系统安全原则、策略,以使不会引发系统安全问题、造成其它危害和引起其它问题(如泄漏用户私密信息);能力可信指的是指软件的广义可靠性、广义安全性及广义正确性.

定义 7 广义可靠性:在规定环境及规定运行时间

内,包括在其它因素的干扰或攻击下,软件引起系统失效的概率及所造成损失的严重程度.

定义 8　广义安全性:在规定环境及规定运行时间内,软件引起系统入侵及造成不可接受风险的概率及所造成损失的严重程度.

定义 9　广义正确性:在规定环境及规定运行时间内,软件产品不能在可接受精确度范围内提供正确或一致结果所造成损失的严重程度.

注:传统可靠性、安全性和正确性的评估度量仅从"正面"考虑,如功能正确率达到多少,容易忽略不正确的功能实施会造成多大损失. 相同正确率的软件,有可能造成的成本损失会有很大差异,这正是传统可信度量容易忽略的. 本文提出的广义可靠性、广义安全性、广义正确性不仅兼容了传统的可靠性、安全性和正确性,更提出了新量纲"损失量",从"反面"考虑软件可信性,不仅提供了新的视角,而且统一了度量量纲,方便将软件可信性不同部分整合形成统一的、可比较的度量评估结果.

定义 10　损失量(Cost-Loss):由于软件存在不可信因素,使得软件使用者需要额外支付的成本(包括金钱、工作量、代码行数等),这些成本由以下三部分构成:由于软件未达到指定行业标准,为使其达到要求而需支付的额外成本;由于不可信问题使得系统无法正常运行,为恢复软件运行而需支付的额外成本;软件的不正常运行给用户的业务工作带来的损失成本.

定义 11　软件信任体系:在特定运行环境 $E(t,r)$ 下($E(t,r)$为时间 t 及关系 r 的函数),用户根据软件可信性需求 STR,借助权威评估机构 AEA,对软件可信性进行评价、判断,若选择:

①信任:则信任方(软件用户)一定会或多或少承担信任的"后果",此即可信风险;

②不信任:则进一步修正 STR,重新对软件可信性进行设计、研发与评估,直至选择①.

软件用户一旦选择信任,即意味着软件投入正式使用,而软件可信风险造成的后果,最终可以用损失量(Cost-Loss)来统一量化. 损失量若超过软件用户可接受的域值,则启动"软件可信动态演化".

定义 12　软件可信动态演化:(如图 3 所示)软件可信性的演化过程有两种:

①从 STR 经 1.1 演化为软件的原始可信性 ST,ST 经 1.2 由 AEA 形成评估报告,评估报告经由 1.3 与 STR 进行对比(或组织专家评审),若评审不通过,则重新启动演化过程①,否则就进入演化过程②;

②一旦启动信任就进入 2.1,而信任后果经由 2.2 将影响 STR 及 ST,其中调整后的 STR 经由 2.3 也将进一步影响 ST,ST 经研发人员重构、修改后经 2.4 由 AEA 再次形成评估报告,评估报告经由 2.5 与新版 STR 进行对比(或组织专家评审),若评审不通过,则重新启动演化过程①,否则就再次进入演化过程②.

如此循环反复,使得 ST 在不断演化的过程中得以提高,直至软件生命周期结束.

6　实验论证

为了验证 STCM 的有效性,我们提出了基于时损率及幂函数的改进 J-M 模型的能力可信度量方法,基于改进的多短板木桶理论的准则可信度量模型及基于同源代码损失量的身份可信度量方法,将软件可信性三个方面的度量通过损失量统一到了一起,因篇幅限制,具体的模型及理论方法请参考文献[20].

基于中国信息安全评测中心(China Information Technology Security Evaluation Center,CNITSEC)及国家信息安全漏洞库的相关数据,我们运用前面提到的方法对多个软件进行了实验和评测,得出的软件可信性结论与实际情况基本一致.

以 Adobe Flash Player 为例,通过某一时间段历史数据,基于时损率及幂函数的改进 J-M 模型,分别计算出广义可靠性系数 $R_f(\bar{x}) = 0.048622$ 及广义安全性系数 $S_v(\bar{x}) = 0.519635$,对照表 7,得到基于损失量的能力可信值 $A(T_a(t)) = 40\text{K}\$$.

表 7　能力可信不同域值损失量对照表

$R_f(\bar{x})$	$A_f(T_a(t))$	$S_v(\bar{x})$	$A_v(T_a(t))$
0 ~ 0.1	10K $	0 ~ 0.1	20 K $
0.1 ~ 0.3	10 K $ ~ 20 K $	0.1 ~ 0.3	20 K $ ~ 50 K $
0.3 ~ 0.5	20 K $ ~ 100 K $	0.3 ~ 0.5	50 K $ ~ 200 K $
0.5 ~ 0.8	100 K $ ~ 200 K $	0.5 ~ 0.8	200 K $ ~ 500 K $
0.8 ~ 1.0	> 200 K $	0.8 ~ 1.0	> 500 K $

同样,基于改进的多短板木桶理论的准则可信度量模型,我们计算得到软件的准则值 7.8,对照表 8,得到基于损失量的准则可信值 $S(T_a(t)) = 50\text{K}\$$.

表 8　大型软件准则损失量对照表

Rule Value	Trustworthiness Level	Loss
0 ~ 2	Level I	1000K $
2 ~ 4	Level II	500K $
4 ~ 6	Level III	200K $
6 ~ 8	Level IV	50K $
8 ~ 10	Level V	10K $

基于同源代码损失量的身份可信指的是引用第三方源代码导致的成本损失可信度量,根据国家信息安全测评中心内部提供的源代码检测工具得到表 9 的引用情况,从而基于损失量的身份可信值 $I(T_a(t)) =$

16.4K $.

表 9 同源代码检测结果对照表

T_i	β_i	α_i	ω_i
Software 1	20%	60%	100K $
Software 2	50%	35%	20K $
Software 3	30%	10%	30K $

基于损失量的统一量纲及不同部分可信性权重,我们可得到软件的整体可信性度量值:

$GT(Ta(t)) = \alpha I(Ta(t)) + \beta S(Ta(t)) + \gamma A(Ta(t)) = 0.1 \times 16.4 + 0.3 \times 50 + 0.6 \times 40 = 40.64$K $.

从评测及实验数据结果,证明我们所提出的软件可信性概念模型是有效的,不仅可度量,还可以为不同软件可信性的比较提供统一的量纲,是软件可信性科学量化新的研究方向.

7 总结

软件"可信性"根源和本质的探讨几十年来一直都没有得到学术界和工业界的充分重视,而科学逻辑实证主义的观点却认为,新式理论必须概念在前,而后才具有可操作性,从而实证检验的结果才能进一步用于验证概念理论.

本文先是形式化说明了当前被广泛引用的软件"可信性"定义彼此间矛盾且不相容,论证了本文研究的重要性和必要性.接着,在社会学信任理论进行大量调研的基础上,构建出基于社会学的信任体系模型 STM 并有所完善,进一步证实从社会学信任理论出发来探讨软件可信性本质问题是正确的、全新的方向.基于 STM,本文根据软件反映出来的特征进行了映射和建模,首次提出基于社会学信任理论的软件可信概念模型 STCM.最后,基于以上研究成果,本文给出了完整的软件可信性概念定义体系,并通过实验论证了模型的有效性,为软件可信性的进一步发展奠定了基础.

然而,本文的研究仍有许多改进之处,比如理论模型的完备性、包容性的论证问题,实证研究反哺证明问题.STCM 定义体系仍然需要在研究进展中补充和完善,未来仍有许多工作要做.

参考文献

[1] 刘克,单志广,王戟,等."可信软件基础研究"重大研究计划综述[M].北京:中国科学基金,2008.145-151.

[2] Department of Defense. Department of defense trusted computer system evaluation criteria[S]. Department of Defense 5200.28-STD,1985.

[3] 汤永新,刘增良.软件可信性度量模型研究进展[J].计算机工程与应用,2010,46(27):12-16.

[4] Sagasti F R,Mitroff I I. Operations research from the viewpoint of general systems theory[J]. OMEGA,1973,1(6):695-709.

[5] Singleton R,Jr Straits B C,Straits M M,et al. Approaches to Social Research[M]. New York:Oxford University Press,1988.

[6] ISO/IEC 14598-1. Information technology-software product evaluation-Part 1:General overview[S]. ISO/IEC,1999.

[7] Rousseau D M,Sitkin S B,Burt R S,et al. Not so different after all:A cross-discipline view of trust[J]. The Academy of Management Review,1998,23(3):1-12.

[8] Schoorman F D,Mayer R C,Davis J H. An integrative model of organizational trust:Past,present,and future[J]. The Academy of Management Review,2007,32(2):344-354.

[9] Mcknight D H,Chervany N L. What trust means in e-commerce customer relationships:An interdisciplinary conceptual typology[J]. International Journal of Electronic Commerce,2001,6(2):35-59.

[10] Mcknight D H,Carter M,Thatcher J B,et al. Trust in a specific technology:An investigation of its components and measures[J]. ACM Transactions on Management Information System,2011,2(2):1-25.

[11] Kee H W,Knox R E. Conceptual and methodological considerations in the study of trust and Suspicion[J]. The Journal of Conflict Resolution,1970,14(3):357-366.

[12] Lewicki R J,Bunker B B. Developing and maintaining trust in work relationships[J]. In Trust in Organizations:Frontiers of Theory and Research,1996:114-139.

[13] Coleman J S. Foundations of Social Theory[M]. Cambridge,MA:Harvard University Press,1990.

[14] Colquitt J A,Brent S A,LePine J A. Trust,trustworthiness,and trust propensity:A meta-analytic test of their unique relationships with risk taking and job performance[J]. Journal of Applied Psychology,2007,92(4):909-927.

[15] ISO/IEC FDIS 9126-1:2001. Information technology-Software product quality-Part 1:Quality Model[S]. ISO/IEC,2001.

[16] ISO/IEC 2382-14:1997. Information technology-Vocabulary-Part 14:Reliability,maintainability and availability[S]. ISO/IEC,1997.

[17] IEEE 610.12:1990. Standard glossary of software engineering terminology[S]. IEEE,1990.

[18] Becker S,et al. Trustworthy software systems:a discussion of basic concepts and terminology[J]. ACM Sigsoft Software Engineering Notes,2006,31(6):1-18.

[19] 王克,邹嘉,戴一奇.基于软件实名认证的系统安全机制[J].清华大学学报(自然科学版),2008,(07):1169-1172.

[20] Yang X, Jabeen G, Luo P, et al. A unified measurement solution of software trustworthiness based on social-to-software framework[J]. Journal of Computer Science And Technology, 2018, 33(3): 603 – 620.

作者简介

杨　曦　女,1977 年 2 月出生,福建福州人. 2005 年云南大学软件工程专业硕士毕业,其后在福州大学从事教学、科研工作. 2014 年考入清华大学博士,研究方向为软件系统与理论、信息安全及软件可信性等.
E-mail: x-yang14@ mails. tsinghua. edu. cn

罗　平　男,1959 年 6 月出生,湖南溆浦人. 清华大学软件学院教授. 研究方向:密码算法设计与密码分析,漏洞分析与攻击,数据库漏洞与安全,木马利用技术,图像安全. 负责和参加国家自然科学基金项目 4 项,国家科技部 973 和“十一五”科技攻关项目 4 项, 部委和学校项目 6 项,横向课题 6 项. 其中作为项目负责人 16 项.